

PhD 13977

DNA AND PROTEIN INTERACTIONS

IN

FILAMENTOUS BACTERIOPHAGES

by

FRANK CSELIK



A dissertation submitted to the University of Cambridge  
in candidature for the degree of Doctor of Philosophy.

Corpus Christi College,  
Cambridge.

September 1985

Családomnak



Abstract

The filamentous bacteriophages are among the simplest of viruses, and may be divided into two Classes on the basis of morphological differences. Their life-cycles involve three assemblies of protein and DNA: the virion itself, a pre-initiation complex and a pre-assembly (g5p.DNA) complex. The overall replication and assembly of virions in infected cells had been thought to decline sharply during stationary phase, but this was shown not to be so. Virus is produced at roughly the same rate by cultures in both log and stationary phase growth.

Methods for purifying and selectively radiolabelling phages were refined. The presence of minor coat proteins (g3p, g6p, g7p, g9p) in the capsids of the Class I phage fd was confirmed, and a corresponding set of minor proteins (at the same relative stoichiometry) was discovered in the Class II phage Pf1. This indicated considerable structural similarity between the two classes of filamentous phage.

Computer prediction of secondary structure suggested that the g3ps have similar signal and membrane insertion sequences. Much structural variation was found between the g6ps and g9ps of the phages, but not the g7ps - which may be involved in the initiation of assembly.

The g5p.DNA complexes of both viruses were isolated in selectively radiolabelled form and shown to be free of minor proteins, which cannot

therefore be pilot proteins directing assembly. The DNA of these complexes was found to be preferentially susceptible to nuclease digestion at residues 5465-5686, a part of the intergenic zone containing, and centred on, the A-hairpin. A short length of DNA was exposed at the other end of the linear complex, though the DNA paths between the ends are of slightly different lengths. The DNA is therefore in a specific linear orientation, comparable with that in the virion.

The three nucleoproteins of the viral life cycle are control points, directing the DNA towards infection (virions), replication (pre-initiation complex) and assembly (pre-assembly complex). It is likely that the exposed hairpins of the DNA, not pilot proteins, control the generation of these structures.

Preface

This dissertation describes work done in the Department of Biochemistry at the University of Cambridge, from October 1982 to September 1985. The dissertation, and all work reported in it, are my own and are not the outcome of collaboration except where otherwise stated in the text. I have not submitted any part of this dissertation for a degree or other qualification at this, or any other, University.

I should like to thank my supervisor, Dr. Richard Perham, for his continuing interest and direction. I would like to thank my colleagues in the Department of Biochemistry, particularly Nicholas J. Short and Dr. Len Packman, for much invaluable advice, both technical and theoretical. My thanks also go to Dr. Don Marvin, for many stimulating discussions, and to Dr. Barbara Bachmann, of the Medical School at Yale, for her kindness and efficiency in supplying bacterial strains. Finally, I am grateful to Katherine V. Boyle and Dr. Julian Mace for checking the proofs of this thesis.

I am indebted to the Medical Research Council for a studentship allowing me to pursue this course of study, and to the College of Corpus Christi and the Blessed Virgin Mary - as to my departmental colleagues - for suffering my continued presence.

Corpus Christi College,  
CAMBRIDGE.

24th September 1985



Frank Cselik

## Contents

vi

Abstract	iii
Preface	v
Contents	vi
Glossary	ix
Chapter 1: INTRODUCTION	
1.1 The Significance of Viruses	1
1.2 General Virology	4
1.3 The Filamentous Bacteriophages	8
1.3.1 Class I capsid structure	10
1.3.2 Capsid proteins	11
1.3.3 Phage DNA	14
1.4 The Structure of Class II Phages	17
1.5 The Life-cycle	20
1.5.1 Adsorption	20
1.5.2 Disassembly	21
1.5.3 DNA replication	24
1.5.4 Protein synthesis	27
1.5.5 The role of g5p	30
1.5.6 Assembly	34
1.6 Perspectives	36
Chapter 2: MATERIALS AND METHODS	
2.1 Materials	38
2.1.1 Chemical reagents	38
2.1.2 Buffers and media	40
2.1.3 Organisms	42
2.2 Plaque Assays	43
2.3 Microbiology	45

2.3.1 Selective radiolabelling of cell cultures	45
2.3.2 Purification of phages	45
2.3.3 Purification of g5p.DNA complexes	46
2.4 Analysis of Proteins	47
2.4.1 Polyacrylamide gel electrophoresis	47
2.4.2 Calibration proteins	48
2.4.3 Silver-staining	49
2.4.4 Fluorography	50
2.5 Reagents for DNA	51
2.6 Deproteinisation of DNA	53
2.7 Agarose Gel Electrophoresis	53
2.8 Dialysis	54
2.9 Scintillation Counting	55
Chapter 3: PREPARATION OF BACTERIOPHAGES	
3.1 Basic Bacteriology and Virology	56
3.2 Phage Biosynthesis	57
3.3 Purification Methods	61
3.4 Protein Analysis	65
Chapter 4: THE PROTEINS OF THE PHAGE CAPSID	
4.1 Selective Radiolabelling of Capsid Proteins	70
4.2 The Minor Proteins of fd Capsids	75
4.3 The Class II Phage Pf1	78
4.4 The Minor Proteins of Pf1 Capsids	79
4.5 The Stoichiometry of the Minor Proteins of Pf1	81
4.6 Discussion	83
Chapter 5: SECONDARY STRUCTURE PREDICTIONS OF VIRAL PROTEINS	
5.1 Protein Secondary Structure	84

5.2 Phage Proteins of Known Structure	87
5.3 The Structures of The g3ps	90
5.4 The Structures of The g6ps	91
5.5 The Pilot Protein Theory	92
5.6 Possible Pilot Proteins - g9p and g7p	93
Chapter 6: THE PROTEINS OF G5P.DNA COMPLEXES	
6.1 The g5p.DNA Complexes	96
6.2 Purification of the fd Complex	98
6.3 Minor Proteins of the fd Complex	101
6.4 The Pf1 g5p.DNA Complex	106
6.4 Discussion	109
Chapter 7: THE DNA OF G5P.DNA COMPLEXES	
7.1 The Possible Orientation of the DNA	111
7.2 Large-scale Preparation of fd g5p.DNA Complexes	112
7.3 Preparation of Replicative Form DNA	114
7.4 Nuclease Digestion of Complexes	116
7.5 The DNase I Susceptibility of Complex DNA	120
Chapter 8: Discussion	
8.1 The Structures of Filamentous Phages	125
8.2 The Nucleoproteins of the Life-cycle	127
8.3 Future Perspectives	130
References	
Appendix I - Useful Figures	
Appendix II - fd and Pf1 Open Reading Frames	
Appendix III - A Listing of Program "Protein Fold"	

Glossary

A <sub>680</sub>	The absorbance at 680 nm, 1 cm pathlength.
AR	Analytical Reagent
ATP	Adenosine Tri-Phosphate
CD	Circular Dichroism
C-terminus	Carboxy-terminus (of a polypeptide).
DNAse I	Deoxyribonuclease I
DTT	Di-Thio-Threitol
EDTA	Ethylene Diamine Tetra-Acetic Acid
Metrizamide	2-[3-Acetimido-5-N-Methylacetimido- 2,4,6-triiodobenzamido]-2-deoxy-D-glucose
moi	multiplicity of infection (number of phages per cell)
$\frac{M}{r}$	Relative molecular mass
MTPA	A defined growth medium
N-terminus	Amino-terminus (of a polypeptide)
pfu	plaque forming units
PMSF	Phenyl-Methyl-Sulphonyl-Fluoride
PPO	Diphenyloxazole
RF	Replicative Form (in DNA replication)
SDS	Sodium Dodecyl Sulphate
(+) strand	The DNA strand relating directly to messenger RNA
(-) strand	The DNA strand complementary to the above
TEMED	N,N,N',N'- Tetra-Methyl-Ethylene-Diamine
TMV	Tobacco Mosaic Virus
Tris	Tris-(hydroxymethyl) aminoethane
2xTY	A rich growth medium

## CHAPTER 1

### INTRODUCTION

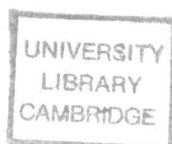


### 1.1 The Significance of Viruses

Viruses, for so long bizarre agents of disease, have acquired a new significance as genetic entities in their own right. Patient, intensive study has shown how their properties may be used to advantage and they are now basic tools in the trade of the molecular biologist.

The medical perspective on viruses is chiefly a combative one; diseases such as typhus and herpes B cause much human suffering. Some, such as rabies, are usually fatal; others, such as influenza, are only moderately discomforting but attack large proportions of the population. A common feature of these diseases is that little can be done to counter an established infection. Only one person has been recorded as surviving the clinical symptoms of rabies, despite sustained efforts at developing an effective therapy. Similarly, aspirin may suppress the symptoms of influenza, but does not combat the virus itself.

The main problem in designing anti-viral agents is that chemotherapy exploits differences between the pathogen and the host; in the case of viruses, much of the life-cycle is executed using the host metabolism. The virus contributes so little to the life-cycle that there is a very limited range of points susceptible to chemotherapeutic attack. Studies of the immune system have demonstrated how it tackles viral infections. The primary response of the body to infections is the production of antibodies to "non-self" antigens. A secondary response has been discovered in the release of interferons, which help contain viral



infections. Biotechnology companies are being geared up to produce interferons for chemotherapy, but the initial promise of a viral panacea has faded somewhat.

These frustrations have been highlighted by the growing realisation of the links between cancer and viruses. Although a tumour may be found without its associated virus, and virus-infected tissue may show no symptoms of malignancy, there is a good correlation between certain combinations of tumours and viruses. More recently, the disease AIDS (acquired immune deficiency syndrome) has been related to a specific virus. A causal connection has not yet been established, for AIDS is not regarded as a valid cause on death certificates. As in the case of cancer, many persons have the virus in their tissue fluids without expressing disease symptoms (about 50% of haemophiliacs receiving factor VIII from human blood).

Viruses have also had considerable impact on agriculture and farming. A variety of stunts, wilts and mottles plague arable farming, while livestock are susceptible to a range of viruses such as fowl plague virus or equine abortion virus. Considerable amounts of gross production are lost each year through the combined effects of these diseases.

On the positive side, viruses have recently become very important in the genetic manipulation of organisms. They are used to target DNA to the host cell, before its stable incorporation into the genetic information of the cell (Ohsumi et al., 1978). The reverse process is

also employed: getting progeny phage to pull out small pieces of host DNA for detailed analysis. Viruses have long been used therapeutically primarily for the inoculation of crippled forms to confer resistance to pathogenic strains. Some interest has been shown in using the infective specificity of viruses as the basis of therapy, but this is an area lacking in development (Smith & Huggins, 1983).

Finally, the simple viruses invite study as reproductive systems which are more amenable to complete description than complex cellular forms. The genetic information of some viruses codes for less than a dozen proteins. It would seem reasonable to start with such simple systems, study how the components inter-relate and how their structures fit them to their functions. This approach to virology has yielded much knowledge and insight into the general mechanisms of biochemical processes. It is to be hoped that in the long term such information will be used to the benefit of mankind in solving the medical and agricultural problems described above.

A concentrated, scientific study of viruses only began when tobacco farmers found their crop being ravaged by a disease which mottled the leaves in a mosaic-like pattern. It was soon realised that the disfiguring principle could pass through filters that excluded even bacteria, but a simple chemical toxin was thought to be responsible for the activity. Physicochemical studies indicated this to be a protein, which was eventually crystallized. Only later was the presence and significance of the nucleic acid realised. Since then, tobacco mosaic virus has been a spearhead in the study of viral construction, assembly and disassembly (Knight, 1974).

It is not surprising that bacterial viruses were discovered last, for their hosts were only recently studied in enough detail for the symptoms of disease to be noticed. However, in a short time they have come to occupy a central place in biochemistry and cell biology. Their study has contributed to the current knowledge of DNA as the genetic material, and demonstrated the role of mRNA in the expression of genomic information. The realisation that viral DNA can be incorporated into the the host cell chromosome lies at the roots of modern genetic engineering, and the use of viruses as vectors of DNA is important in the present day. The causative viruses of many diseases have been isolated, and can now be propagated and studied in the laboratory under controlled conditions.

### 1.2 General Virology

Viruses, as a class of organisms, have many broad features in common. Perhaps the most fundamental is that they are all units of nucleic acid, protected in capsids, using the metabolism of host cells to effect multiplication. DNA and RNA viruses are known, but both polynucleotides are never found in the same virus. The nucleic acid may be either double-stranded or single-stranded, and it may be a linear molecule or covalently closed as a loop. This nucleic acid is the essential virulent agent, for in some cases it is the only part of the virus to enter the cytoplasm of the target cell. Furthermore, if isolated viral nucleic acid is manipulated into the cytoplasm of a suitable host a normal infection may proceed.

Between infections the genome is usually protected, by a surrounding capsid, from nucleases and mechanical damage. This may, in a simple virus, comprise only protein but more complicated viruses include glycoprotein and membrane components. A problem arises in capsid construction: if the capsid structure is defined uniquely at all points, the nucleic acid needed to define that giant protein structure would be much longer than could be accommodated inside it (Horne & Wildy, 1961). Thus virus capsids are constructed from a series of identical subunits, in equivalent relationships. It may be argued that ribosomes are examples of totally asymmetric nucleoprotein aggregates. However the proteins of this type of aggregate are coded in a large quantity of DNA resident in the nucleus, not present in the ribosome. The system is therefore not truly analogous.

A single small coding sequence of viral nucleic acid generally defines a protein which is made in thousands of copies to cover the entire genome. As a result capsids exhibit strong symmetries. A secondary consequence is that such well-defined structures impose a size limit on the nucleic acid they enclose. Most viruses have a genome of defined, near maximal size and cannot accommodate large genomic insertions (Caspar & Klug, 1962; Katsura 1984). Does the capsid define genome size, or the genome define a capsid sufficient to contain it ?

The symmetry of viruses has provoked much theoretical investigation (Caspar & Klug, 1962; Holmes, 1982; Klug, 1983). Three main classes have been identified: icosahedral, helical and complex. In each case the

symmetry need not be based on the subunits constituting the capsid, but may be a symmetry of groups of subunits. For example, geodesic domes are at first sight regular arrays of equilateral triangles. Such an array forms a flat sheet; the dome curves in two dimensions because a triangle is periodically omitted and the true symmetry involves pentameric triangular groups. The triangles are not all in identical environments, so are not equivalent, but they are validly involved in generating icosahedral symmetry and are regarded as "quasi-equivalent" (Caspar & Klug, 1962). This is a crucial aspect of viral architecture. Strict icosahedral symmetry (Fig. 1.1) can accommodate only 60 identical subunits, which puts an upper limit on capsid (and hence genome) size. Quasi-equivalent subunits allow the limit to be extended considerably, and most "icosahedral" viruses actually involve considerable quasi-equivalence.

The symmetry of helical viruses needs little explanation, except to point out that more than one helix may be involved. A given array of subunits may have several theoretical helices drawn through it, each of different pitch. Usually the lowest order helix is assigned, though there may be several interwound helices involved - as in the double-stranded structure of DNA or the triple helix of collagen fibres.

The complex viruses show a composition of different symmetry types in one capsid particle. T-even phages have a "head" region, enclosing the DNA, of icosahedral symmetry which joins a stalk and baseplate of 6-fold symmetry. The complex viruses form a continuous spectrum of composite

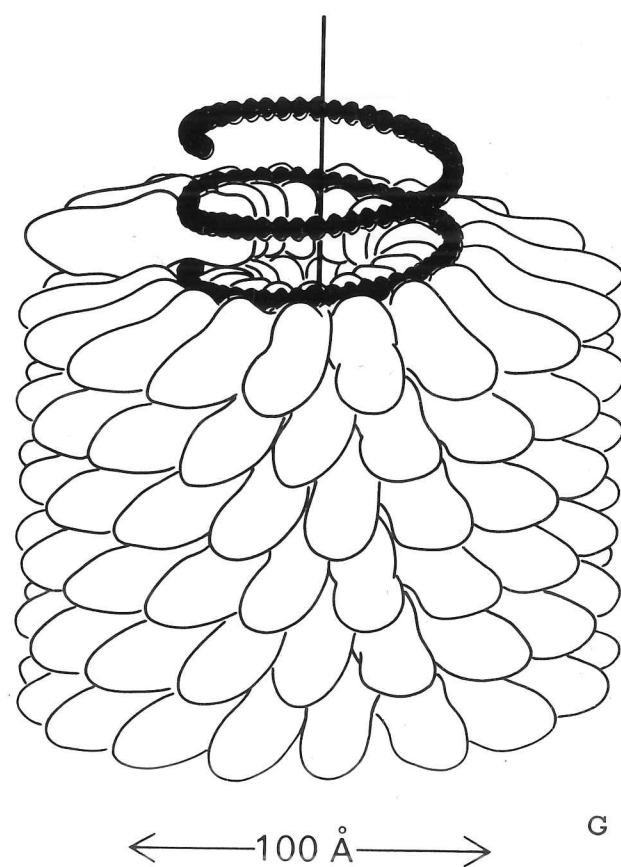
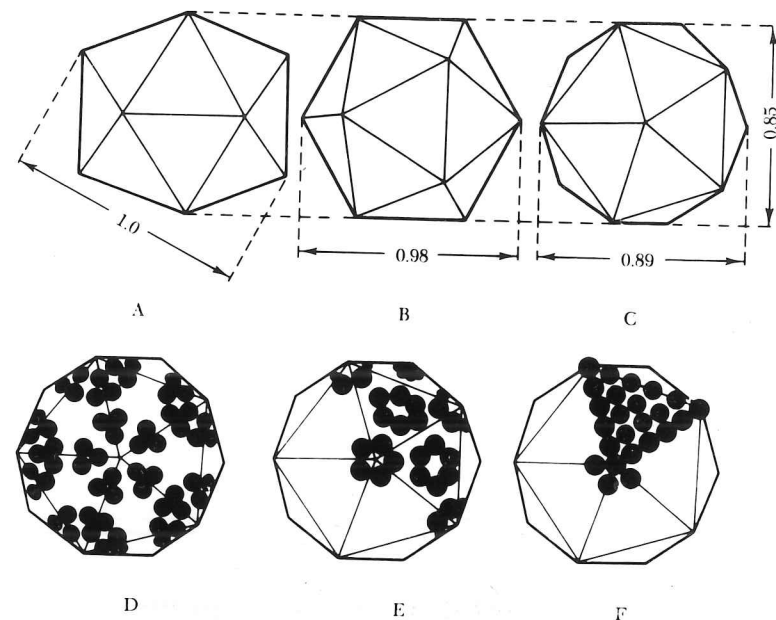


Figure 1.1. The symmetries of virus particles. Icosahedral symmetry (views a-c) gives an isodiametric shell of protein around the DNA. The protein subunits can be arranged in a variety of ways within such a symmetrical array (d-f). Helical symmetry (g) gives an elongated, filamentous particle; note that the symmetry is broken at the two ends of such an array. Complex symmetry combines these two symmetries in one particle, with an icosahedral "head" and helical stalk.



symmetries, but the basic principles of symmetrical construction apply to each of their parts considered separately (Kornberg, 1974). Of course, the junctions between these symmetries are of considerable interest in their own right (Kochan et al., 1984).

Many animal viruses also bear an outer membrane derived from the plasma membrane of the host cell from which they were released. Usually, these membranes contain virus-coded proteins in addition to the normal host cell proteins. The membrane does not significantly increase the protection of the nucleic acid, but is thought to fuse with the target host cell membrane to facilitate penetration and infection. This process of finding a suitable host is specific to an extent depending on the virus under consideration. Generally, plant and animal viruses have a wider host range than bacterial viruses, which may infect only particular strains of a species of bacterium (Luria et al., 1978).

Once the virus has adsorbed to and entered the host, viral nucleic acid is released to the cytoplasm. The biosynthetic machinery of the host is used to replicate the DNA and produce viral proteins, including those of the capsid. Later the progeny nucleic acid and capsid proteins come together to assemble mature virions. The assembly process is thought to be driven by interactions between the subunits themselves in a spontaneous process of self-assembly (Caspar & Klug, 1962). Such self-assembly was first demonstrated in vitro for tobacco mosaic virus, using purified viral RNA and capsid protein which associate, under suitable conditions, to give mature virions (Fraenkel-Conrat & Williams, 1955).



In a lytic life-cycle the assembled progeny build up inside the host cell until it bursts. In lysogenic life-cycles the sequence of events is interrupted after penetration. The viral nucleic acid is spliced into the host chromosome and remains genetically dormant. Thus it is replicated with the chromosome giving rise to a whole clone of cells, each with a latent viral genome. Once the cell becomes stressed or senescent, the viral DNA is excised by the action of a derepressed viral enzyme. A normal lytic life-cycle follows, leading to progeny phage and cell lysis. Overall, the virus allows itself to be replicated slowly when times are good, and turns on its host when conditions are less favourable (Luria et al., 1978).

### 1.3 The Filamentous Bacteriophages

A typical filamentous bacteriophage has a flexible, cylindrical capsid of protein about 1000 nm long and 60 nm in diameter (Fig. 1.2). This encloses a single closed loop of single-stranded DNA (about 6,000 bases) which has ten open reading frames. These viruses are specific to hosts containing a pilus-producing plasmid; the first be discovered was phage f1, isolated from sewage as a virus specific for F<sup>+</sup> strains of the Gram-negative bacterium Escherichia coli (Loeb, 1960). Several further strains have been isolated including fd (Marvin & Hoffman-Berling, 1963), M13 (Hofscneider & Preuss, 1963), ZJ/2 (Bradley, 1964), Ec9, If1, Ike, Xf and v6 (reviewed Marvin & Hohn, 1969). The strains fd, f1 and M13 are very closely related, having no more than 3% difference between their genomes, and may be regarded as strains of the same bacteriophage,

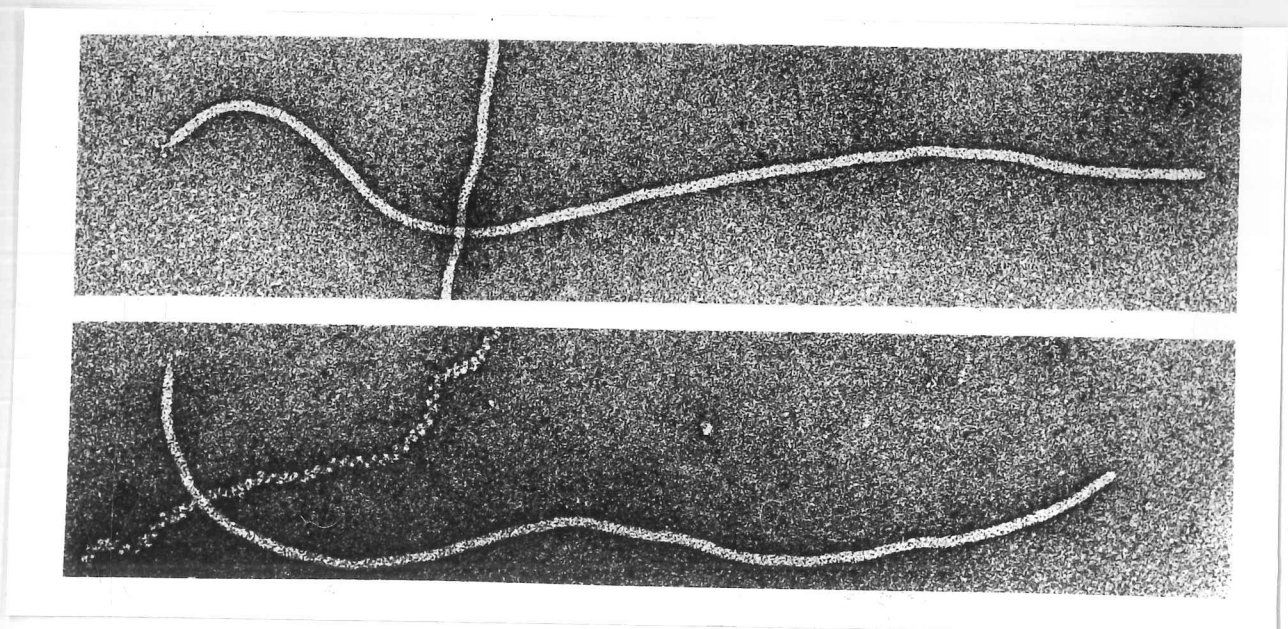


Figure 1.2. Electron micrographs of the filamentous bacteriophage fd. The virus has a long, flexible capsid of helically arranged subunits. Some g5p.DNA complex is included for comparison in the lower picture (section 1.5.5). At one end of the phage particle are a few small globular structures, the g3p. The bar represents 100 nm.

(After Gray et al., 1981.)

termed Ff (Hill & Petersen, 1982). The discovery of the phages Pf1 and Pf3, infecting strains of Pseudomonas aeruginosum, led to the division of the filamentous group into two types: Class I, described above; and Class II, having capsids about twice as long with a lower proportion of DNA to protein. Pf1 is the archetype of Class II, and the most extensively studied of that class.

Filamentous viruses may be studied from several different viewpoints. The major capsid protein provides an excellent model for the study of membrane proteins, their biosynthesis and membrane insertion (Wickner, 1983). The small genome has made convenient the study of DNA replication and the control of gene expression (Marvin & Hohn, 1969). The lack of stringency in genome size allows the phages to be used as cloning vectors (Heurmann et al., 1978; Ohsumi et al., 1978; Schaller, 1978), and the large number of mutants available in both the viruses and their hosts facilitates investigation of the genetics of these systems (Pratt et al., 1966). Yet perhaps the most intriguing study of all is that of the viral life-cycle at the molecular level. These endoparasites manage to preserve their genomes outside their hosts, recognise and infect the host, reproduce and disseminate the progeny though coding only for ten of their own proteins - many of which are fairly small. Study of these highly adapted proteins and their interactions may give greater insight into the field of protein structure in relation to function, and aid in the development of a "systems" approach to protein biochemistry.

### 1.3.1 Class I Capsid Structure

The filamentous nature of bacteriophage fd was proven by the isolation of infectivity specific for  $F^+$  strains of E.coli which contained fibrous material as shown by electron microscopy (Marvin & Hoffman-Berling, 1963; Zinder et al., 1963). This was novel, as the only single-stranded DNA virus then known ( $\Phi$ X174) was of the icosahedral morphology. Filament length has been derived from calibrated electron microscopy as  $880 \pm 20$  nm and from diffusion coefficients as  $895 \pm 20$  nm (Newman et al., 1977; Frank & Day, 1970). The latter also gave a value of  $9 \pm 1$  nm for the diameter which is large in comparison with the value from X-ray diffraction studies: 5.5 nm (Wachtel et al., 1974). The relative molecular mass of Ff phage particles has been determined as  $M_r$   $16.4 (\pm 0.6) \times 10^6$ , the DNA content (6408 nucleotides in fd) comprising 12% of the virion by mass (Newman et al., 1977).

From the earliest papers in this field it was suspected that the DNA was enclosed within a tubular protein sheath, though there was some controversy over whether the DNA molecule passes up the length of the virion in one protein sheath and down in a second, parallel one or has a single sheath encompassing both lengths of the DNA loop (Hoffman-Berling et al., 1966; Marvin, 1966). The axial structure of the virion was solved from the X-ray diffraction measurements which showed clearly that the capsid is a single, hollow tube of protein enclosing the DNA (Wachtel et al., 1974; Banner et al., 1981). The unit cell displays an electron-dense core up to 1.3 nm on the radius - the DNA - surrounded by a single shell of lower electron density - the protein - between 2 and 3 nm radius.

The nature of the helical symmetry in the protein capsid has also been a matter of disagreement, owing to differing interpretations of X-ray diffraction patterns. A single start helix was proposed with 22 subunits in 5 turns, and a regular perturbation which demanded considerable quasi-equivalence. This has now been supplanted by the model of a five start helix, with a rise between successive subunits of  $0.319 \pm 0.015$  nm (Caspar & Makowski, 1981; Newman *et al.*, 1977; Wiseman & Day, 1977; Wachtel *et al.*, 1974). The capsid may be formed notionally by the stacking of pentameric rings of subunits, each ring rotated  $36^\circ$  with respect to the one below it.

### 1.3.2 Capsid Proteins

The viral capsid is formed largely from the product of gene 8. The amino-acid sequence of this protein has been determined by a variety of methods giving a consensus sequence of 50 residues,  $M_r$  5169 and an estimated  $2710 \pm 10$  subunits per capsid (Bailey *et al.*, 1977; Newman *et al.*, 1977; Sugimoto *et al.*, 1977; Nakashima & Konigsberg, 1974; Asbeck *et al.*, 1969). The high proportion of alanine, and application of the rules of Chou & Fasman (1974), indicate considerable alpha-helical structure in the native protein (Marvin, 1966). The ORD spectrum of phage confirms this view, assigning most of the structure as alpha-helix (Nozaki *et al.*, 1976; Day, 1966). Model building studies show how these rods of helix may be arranged in the capsid to give the known helical symmetry. The subunits are thought to overlap like the scales of a fish, with the basic C-terminal end of each forming a lining on the inside of

the capsid, and the acidic N-terminal region exposed at the outer surface (Cross & Opella, 1985; Marvin & Wachtel, 1975). The subunits are held in a fairly rigid array, no large scale motions being detectable in NMR, though a high frequency vibration appears to be present at around  $10^9$  Hz (Cross & Opella, 1982).

The alpha-helical rod is not the only known conformation of g8p. Five states have been proposed, stable under different conditions (Nozaki et al., 1978). These are:

- 1: The 100% alpha-helix - as found in virions;
- 2: The 50% alpha-helix, 50% beta-pleat - as found in membranes;
- 3: The 100% beta-pleat - in 5M guanidinium chloride (GuCl);
- 4: The 50% beta-pleat, 50% random coil - aggregates in 6M GuCl;
- 5: The 100% random coil - in 8M GuCl.

(Williams & Dunker, 1977; Cavalieri et al., 1976; Nozaki et al., 1976; Makino et al., 1975; Chun et al., 1974). The mixed alpha-helical / beta-pleat form appears to be the most stable, and is the form to which the 100% alpha-helical conformer may be converted by gentle denaturation at a hydrophobic surface (such as a membrane). The reverse of this process has only been observed during bacteriophage assembly, an inherently complex process (Griffith et al., 1981).

Minor protein components have also been discovered in the capsid, as first suspected by Henry & Pratt (1969). The g3p is present in probably 5 copies per virion (Lin et al., 1980; Goldsmith & Konigsberg, 1977; Rossomando & Millstein, 1971). Each has a large globular N-terminal

the capsid, and the acidic N-terminal region exposed at the outer surface (Cross & Opella, 1985; Marvin & Wachtel, 1975). The subunits are held in a fairly rigid array, no large scale motions being detectable in NMR, though a high frequency vibration appears to be present at around  $10^9$  Hz (Cross & Opella, 1982).

The alpha-helical rod is not the only known conformation of g8p. Five states have been proposed, stable under different conditions (Nozaki et al., 1978). These are:

- 1: The 100% alpha-helix - as found in virions;
- 2: The 50% alpha-helix, 50% beta-pleat - as found in membranes;
- 3: The 100% beta-pleat - in 5M guanidinium chloride (GuCl);
- 4: The 50% beta-pleat, 50% random coil - aggregates in 6M GuCl;
- 5: The 100% random coil - in 8M GuCl.

(Williams & Dunker, 1977; Cavalieri et al., 1976; Nozaki et al., 1976; Makino et al., 1975; Chun et al., 1974). The mixed alpha-helical / beta-pleat form appears to be the most stable, and is the form to which the 100% alpha-helical conformer may be converted by gentle denaturation at a hydrophobic surface (such as a membrane). The reverse of this process has only been observed during bacteriophage assembly, an inherently complex process (Griffith et al., 1981).

Minor protein components have also been discovered in the capsid, as first suspected by Henry & Pratt (1969). The g3p is present in probably 5 copies per virion (Lin et al., 1980; Goldsmith & Konigsberg, 1977; Rossomando & Millstein, 1971). Each has a large globular N-terminal



domain ( $M_r$  37,600), released by subtilisin digestion, and a small (presumably fibrous) domain attaching it to the virus particle (Armstrong *et al.*, 1981). This structure has been observed in electron micrographs of intact phages (Gray *et al.*, 1981). Three other minor proteins have been reported in the capsids of Ff phages, termed C, C' and D. These are encoded by genes 9, 7 and 6 respectively and appear to be present in 5 copies each per virion (Simons *et al.*, 1981; Lin *et al.*, 1980; Simons *et al.*, 1979). Fragmentation of phage followed by immuno-selection of the g3p end of the capsid showed that g6p is present at that end whereas g7p and g9p are present at the opposite end (Fig. 1.3), (Grant *et al.*, 1981). It would seem reasonable to rename C'-protein (g7p) as E-protein, but this has not yet taken place in the literature.

The discovery of such minor protein components has given rise to much speculation about their functions. In particular g3p has been regarded as a pilot protein, guiding the DNA through the life-cycle, whereas g8p is merely a structural feature, protecting the DNA between infections (Kornberg, 1982; Kornberg, 1974). The g3p was thus proposed not only to cause specific attachment to host cells, but to aid the penetration of DNA to the cytoplasm, direct its replication while anchoring it to the membrane and initiate the process of assembling mature virions. It was also suggested that each function is associated with a different face or active site of the pilot protein molecule.



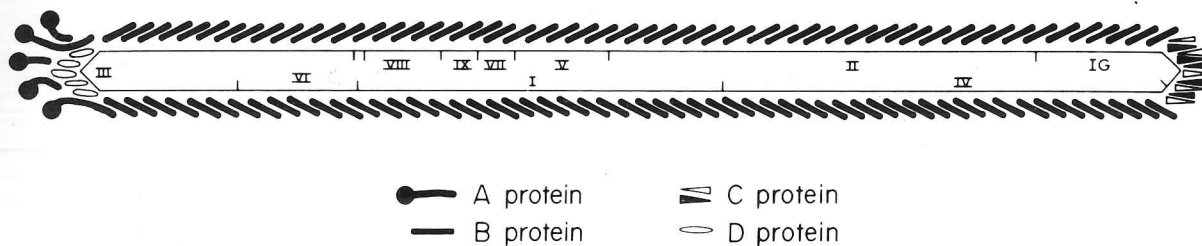


Figure 1.3. Diagram of the structure of the fd capsid. The axial loop of single-stranded DNA is enclosed in a helical array of g8p (B-protein) subunits. At one end of the filament are located five copies of g3p (A-protein), the pilus-specific adsorbing moiety. The same end bears five copies of g6p (D-protein), and the opposite end bears five each of g7p and g9p (C-protein). The DNA is oriented with the intergenic zone (IG) at the g7p.g9p (C-protein) end of the capsid. No corresponding data are available for phage Pf1.

(After Webster et al., 1981.)

The evidence from experiment, however, does not support the role of g3p as a pilot protein. Any specific interaction of the g3p with the viral DNA is unlikely, since the DNA sequence behind this protein in the virion can be replaced with DNA from other viruses, by genetic manipulation, without affecting viability. Possible further roles of these minor proteins are obscure, beyond the knowledge that strains with lesions in these genes cannot assemble mature virions despite making g8p and DNA as normal.

### 1.3.3 Phage DNA

Both classes of filamentous bacteriophages contain single-stranded DNA. This is a somewhat unusual form for the genetic material, which is more often double-stranded with the bases hydrogen bonded to each other. The strand is not extensively base-paired for the base composition shows non-Watson/Crick ratios of A:T and C:G, no thermal hyperchromicity is observed and antibodies to double-stranded DNA fail to bind (Thomas & Prescott, 1983; Levy & Simpson, 1973; Yamamoto & Naito, 1965; Salivar et al., 1964). The DNA is resistant to exonucleases, indicating it to be one covalently closed loop (Marvin & Schaller, 1966); complementation analysis confirms that the genome is circular (Fig. 1.4), (Lyons & Zinder, 1972; Pratt et al., 1966).

After considerable effort (Hill & Petersen 1978; Bailey et al., 1975; Tate & Petersen, 1974a,b) the complete DNA sequence of Ff phage has been obtained, mainly by Maxam-Gilbert methods (Hill & Petersen,

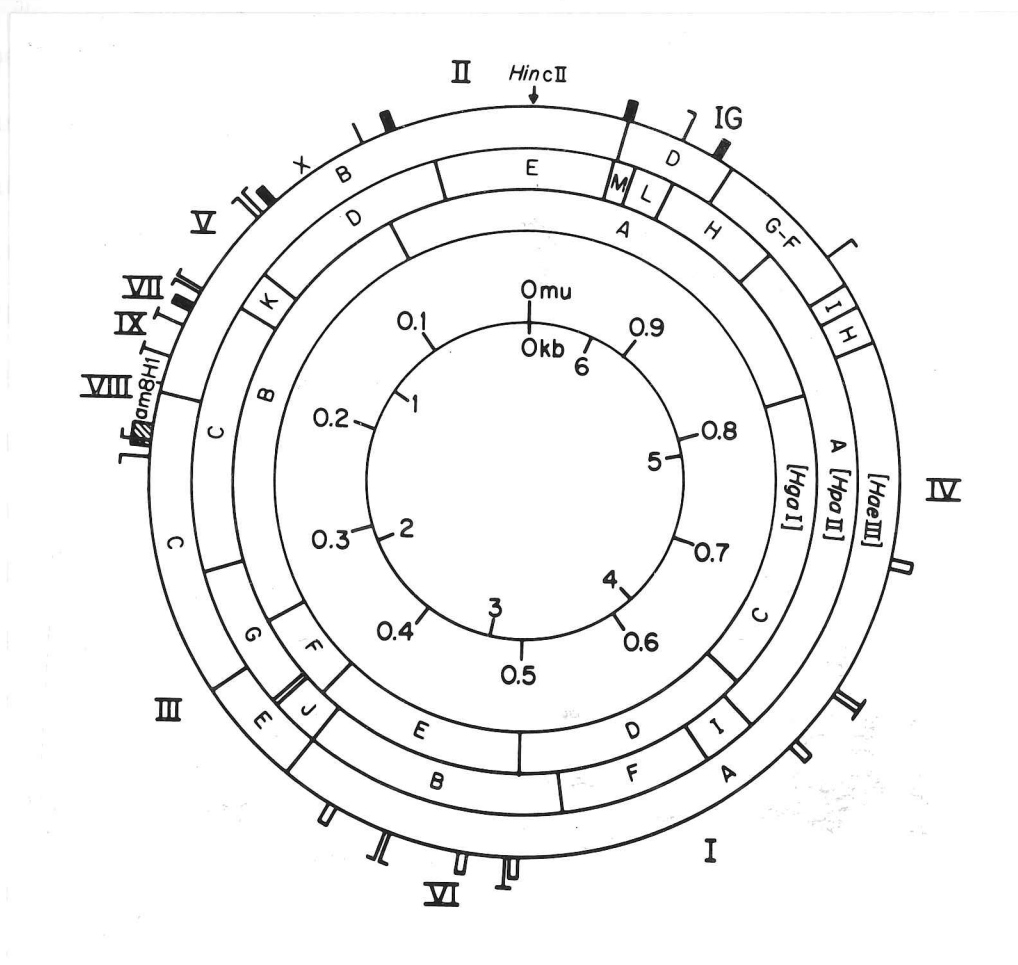


Figure 1.4. Genetic and physical map of bacteriophage fd DNA. The outer circle represents the arrangement of genes I to IX on the circular DNA, and the position of the intergenic zone (IG). G-start promoters of transcription are represented by solid boxes, A-start promoters by open boxes, and the single rho-independent terminator by a shaded box. The central scale is marked in map units (mu) and kilobases (kb). Cleavage maps of three restriction endonucleases are shown, together with the single cleavage site of Hinc II.

(After Moses et al., 1980.)

1982; Beck et al., 1978). The fd genome contains 6408 nucleotides and carries ten open reading frames with only one short overlap and a large intergenic space (Beck et al., 1978; Van den Hondel et al., 1976). This lies between genes 4 and 2, and contains several palindromic sequences (Fig. 1.5) capable of base-pairing to form hairpins (Webster et al., 1981). Polypeptide products may be predicted from the DNA sequence, which may correlate with functions known from the genetic map. An open reading frame cannot, however, be taken as definitive evidence of the existence of a functional protein. Gene 10, for example, may be a significant coding sequence but this has not been resolved either genetically or by protein analysis.

The conformation of the viral DNA has not been unequivocally determined. Free of the capsid, it has been visualised by electron microscopy as a circular loop of contour length 2000 nm. Thus it must be present as a pair of antiparallel strands in the 1000 nm capsid (Inman, 1967). The encapsidated bases appear to be extensively base-stacked and inclined to the axis with a semi-angle of  $25^\circ \pm 5^\circ$ , shown by spectroscopic investigations (Day, 1969; Bendet & Mayfield, 1967). One model suggests an extended conformation with the phosphate residues nearer the central axis than the bases, another that the DNA is in a pseudo-base paired double helix (Day et al., 1979).

The problem of DNA conformation is very relevant to that of how the DNA interacts with the capsid proteins. In TMV, for example, the RNA has a precise stoichiometry of 3 nucleotides per protein subunit, each



binding site being identical (Butler, 1984). In the Ff phages the ratio of nucleotides to subunits is estimated as  $2.3 \pm 0.11$  (Newman *et al.*, 1976),  $2.34 \pm 0.15$  and  $2.34 \pm 0.06$  (Berkowitz & Day, 1977). None of these allows the possibility of an integral number, so a model of g8p as a DNA binding protein which aggregates into tubular capsids (like TMV) is excluded. Any interaction of the DNA with the capsid must be non-specific, such as general electrostatic interaction between the basic amino-acids lining the capsid and the phosphates of the DNA (Cross & Opella, 1985; Berkowitz & Day, 1976; Chun *et al.*, 1974).

It was once thought that the DNA had a random orientation within the capsid (Tate & Petersen, 1974). This has now been shown not to be the case (Webster *et al.*, 1981; Shen *et al.*, 1979). A specific region of DNA, at the intergenic space, is not found randomly located in different capsids, but always at the g7p.g9p-end of the virion. Genetic manipulation can give rise to strains with extra DNA inserted into the genome, shifting the DNA round in the capsid, but the inter-genic zone is always found at the g7p.g9p-end of the capsid. Any sequence may be present at the g3p.g6p-end (Ikoku & Hearst, 1981).

It has already been pointed out that the intergenic zone contains several palindromic sequences capable of forming double-stranded hairpins (Fig. 1.5). A series of these from A to E has been located by computerised folding methods (Rowe, 1983; Webster *et al.*, 1981), and the

presence of some of them has been confirmed experimentally (Huang & Hearst, 1981; Niyogi & Mitra, 1977; Sishido & Ikeda, 1977). Crosslinking of the DNA strands in phages shows that a double-stranded loop exists in the particle, its tip corresponding roughly with the position of hairpin A (Ikoku & Hearst, 1981; Webster et al., 1981).

In summary, the current view of Ff phage architecture, at the time this work was commenced, was that of a tubular capsid of g8p, as a five start helix, enclosing the circular DNA loop without specific interactions. The DNA has a hairpin in the intergenic zone which always resides at the g7p.g9p-end of the virion. The g3p.g6p-end is responsible for host infection and has no sequence specific interaction with the DNA.

#### 1.4 The Structure of Class II Phages

The most studied member of Class II is the filamentous phage Pf1, infecting Pseudomonas aeruginosum K (Takeya & Amako, 1966), which is compared with fd in Table 1.1. In common with the Class I phages the capsid is an elongated tube, but it is much longer than that of fd, around 1960 nm, and contains  $7620 \pm 440$  subunits (Wiseman & Day, 1977). This has proven more amenable to investigation by X-ray diffraction than the Ff type, showing an array of subunits as a single start helix with a pitch of 1.5 nm and 27 subunits in each 7.5 nm structural repeat

Table 1.1. A Comparison of Phages fd and Pf1

Bacteriophage	fd	Pf1
Class	I	II
Host	<u>Escherichia</u>	<u>Pseudomonas</u>
	<u>coli</u>	<u>aeruginosum</u>
Length	880 nm <sup>a</sup>	1960 nm <sup>b</sup>
DNA size (nucleotides)	6408 <sup>c</sup>	c. 7390 <sup>d</sup>
% DNA in virion <sup>d</sup>	12	6
Subunits per virion <sup>d</sup>	2,710 ± 110	7,620 ± 440
Bases DNA per subunit <sup>d</sup>	2.32 ± 0.07	0.97 ± 0.05
Capsid helix <sup>e</sup>	5 start	1 start
Helix pitch <sup>e</sup>	16 nm	1.5 nm

From: a) Frank & Day (1970); b) Marvin & Hohn (1969); c) Beck et al. (1978); d) Day & Wiseman (1978) e) Caspar & Makowski (1981).



(Nave et al., 1981; Wiseman & Day, 1977; Marvin & Wachtel, 1975; Marvin et al., 1974a). At room temperature the capsid shows 5.4 subunits per turn whereas at 4°C the figure becomes 5.46 (Thomas & Prescott, 1983; Marvin et al., 1981; Hinz et al., 1980; Nave et al., 1979; Wachtel et al., 1976). Accumulated over the length of the virion, this torsion gives a total of 15 rotations of one end of the phage with respect to the other and a reduction of 100 nm in length (Torbet, 1979). This is a discrete, co-operative transition with a free energy change of  $14.5 \text{ kJ.mol}^{-1}$  in g8p and the co-operative unit is 26 subunits, similar in size to the structural repeat. Such a conformational transition may precede disassembly (Marvin et al., 1981).

A major capsid protein is present - g8p - the 46-residue sequence of which is known (Wiseman et al., 1976; Nakashima et al., 1975; Nakashima et al., 1974). Correlation of Chou-Fasman analysis of the sequence with the data from X-ray diffraction led to the suggestion of a two-domain coat protein. Residues 1-22 and 26-46 are in alpha-helix, the linker being in an extended conformation. The domains form two rings of high electron density in the capsid, at 1.5 nm and 2.5 nm radius, the inner sheath being more tightly packed than the outer one (Makowski et al., 1980; Nakashima & Konigsberg, 1980).

The phage DNA is a single-stranded loop of about  $7390 \pm 180$  nucleotides which resides in the capsid largely as two antiparallel

strands (Wiseman et al., 1976). Since the ratio of nucleotides to subunits is approximately 1:1, there may be fundamentally different interactions between the DNA and capsid compared with the Class I phages. There may even be a specific binding interaction per nucleotide, analogous to the case in TMV. This has led to speculation over the degree of evolutionary relationship between the two classes of bacteriophage. The question arises: are the classes the result of divergence from a common filamentous ancestor or are they derived from nucleoprotein complexes which happened to have similar symmetries and then evolved towards the same form? Possible divergence, based on differences in the coat protein sequences, is estimated at  $3 \times 10^9$  years before present - a long time before the earth formed (Nakashima et al., 1975; Nakashima et al., 1974).

The nature of the binding interactions is also of interest, but remains obscure, as do many details of the life-cycle. Does the capsid contain minor proteins, supporting divergent evolution, or are they absent? In general, ignorance of many aspects of the Class II phages has arisen as a result of the intensive study of the Ff types (Class I) without corresponding investigation of the Class II examples. In consequence the potentially rich methodology of comparing and contrasting functionally related systems has seen limited application thus far.

### 1.5 The Life-Cycle

In common with most other viruses, the filamentous phages show a life-cycle which is divisible into five main phases:

1. Specific attachment of the virus to its host cell;
2. Penetration of DNA to the cytoplasm;
3. Replication of DNA and protein synthesis;
4. Assembly of virions, often via precursors;
5. Release of the progeny from the cell.

These will now be discussed individually.

#### 1.5.1 Adsorption

The host range of filamentous phages is extremely narrow compared with, for example, plant viruses. The reason behind this seems to be that plant viruses enter cells through lesions, whereas the filamentous phages bind specifically to plasmid induced receptors on the host cell. The Ff phages infect only cells expressing the F-plasmid (Loeb, 1960). The virus adsorbs at the tip of the F-pilus elaborated by such strains, as observed by electron microscopy (Fig. 1.6), (Caro & Schnos, 1966; Brinton et al., 1964).

The F-pilus is a long cylindrical protein oligomer anchored in the cell membrane at regions where the inner (plasma) and outer cell membranes meet. It is composed of a single protein species, pilin ( $M_r$  11,200) in a four-fold helical array with a pitch of 12.8 nm and

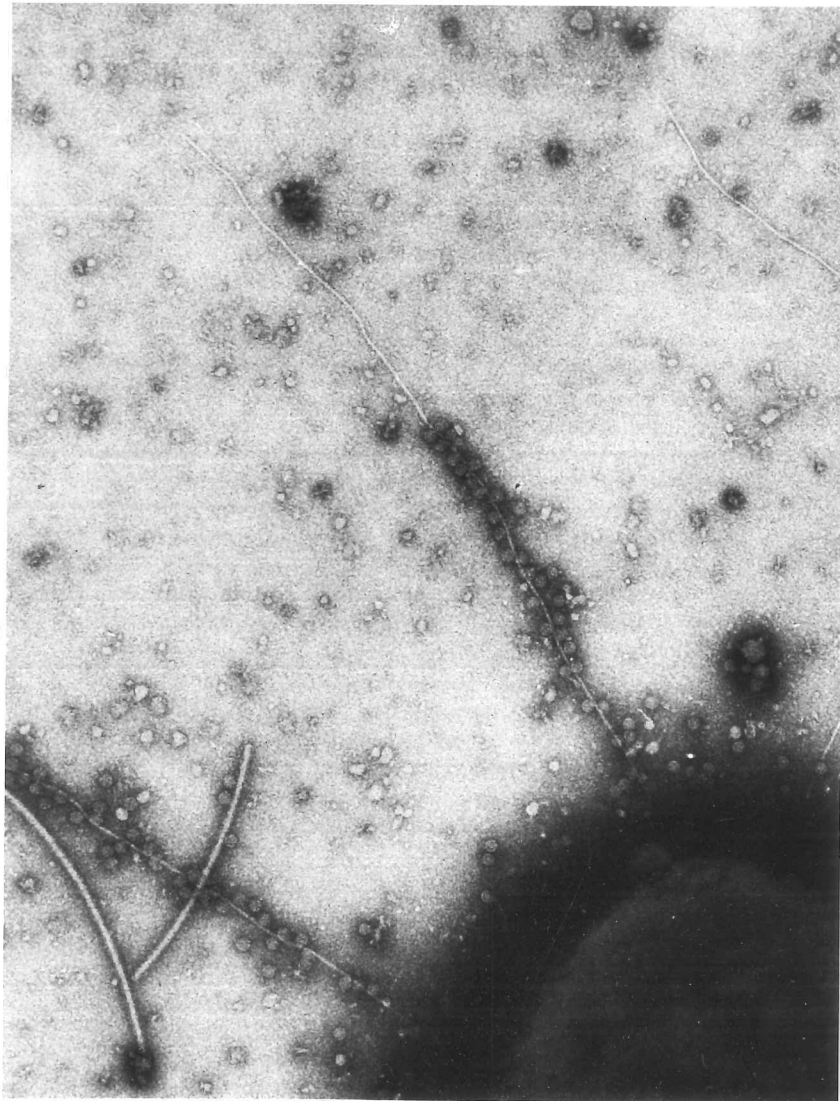


Figure 1.6. The adsorption of bacteriophage fd. In the electron micrograph the slender viral filament is shown adsorbed to the tip of an F-pilus on an E.coli cell. The pilus is decorated with the icosahedral virus MS2, which binds to the sides of the pili but not to their tips. The bar represents 500 nm.

(From Caro & Schnos, 1966.)

diameter 8nm (Folkhard et al., 1979). The pilus is necessary for bacterial conjugation, attaching to receptors in the membranes of F<sup>-</sup> cells (Novotny et al., 1969; Anderson, 1958). Ff phages compete with these receptors for the pilus tip, inhibiting the mating process (Marco et al., 1974; Novotny et al., 1968). The globular domain of g3p is the adsorbing moiety, for when isolated it acts competitively to inhibit Ff phage adsorption (Armstrong et al., 1981). These competition experiments suggest a 1:1 binding stoichiometry of virion to pilus (Fareed et al., 1966). Binding is not dependent on the energy status of the cell, proceeding at 0°C with cyanide-inhibited cells, though the number of pili per cell is reduced (Yamamoto et al., 1980; Tzagoloff & Pratt, 1964).

#### 1.5.2 Disassembly

Early in the course of filamentous phage research, it was thought that the DNA was injected into the host cell, through the axial hole in the F-pilus, which performed its normal function in conducting DNA. Marvin (among others) suggested that the pilus retracted, serving only to bring the virion to the surface of the membrane and thus the site of disassembly (Marvin & Hohn, 1969). The latter theory was supported by electron microscopical investigation of the timecourse of infection: viral DNA only appears in the cytoplasm after pilus retraction (Novotny & Fives-Taylor, 1974; O'Callaghan et al., 1973; Jacobson, 1972). The process of pilus retraction is energy-dependent, being inhibited at low temperatures or by potassium cyanide. Some evidence suggests that the

extension and retraction of pili is a regular, cyclic process. A phage adsorbs onto a pilus and simply waits until the pilus retracts (Novotny & Lavin, 1971; Novotny et al., 1969; Tzagoloff & Pratt, 1964). Once the phage has come up to the membrane, the capsid disaggregates, the subunits inserting into the membrane as the DNA is released on the cytoplasmic side. How the highly polar DNA penetrates the hydrophobic membrane interior is as yet unknown.

Disassembly has been followed in vivo by fractionating cells infected with phages containing radiolabelled coat proteins, showing that the label is present as g8p in the membrane fraction. Proteolytic experiments showed that the inserted g8p spans the membrane only once, N-terminus outside, with residues 1-20 exposed, 21-39 in the membrane and 40-50 in the cytoplasm (Ohkawa & Webster, 1981; Chamberlain et al., 1978; Williams & Dunker, 1977). This arrangement reflects the known primary sequence (Bailey et al., 1977), exposed residues being predominantly polar and those in the membrane nonpolar (Fig. 1.7). The negatively charged end of the protein is outside the cell, and the positive end inside, so insertion proceeds down the electrochemical gradient of the membrane.

Attempts to reproduce this insertion in vitro have highlighted many aspects of viral disassembly and insertion of proteins into membranes. Alkaline disaggregation of phage is a co-operative process, beginning with the ionization of a phenolic hydroxyl of tyrosine (Dunker et al., 1979; Frank & Day, 1970; Rossomando & Zinder, 1968). Similar forced

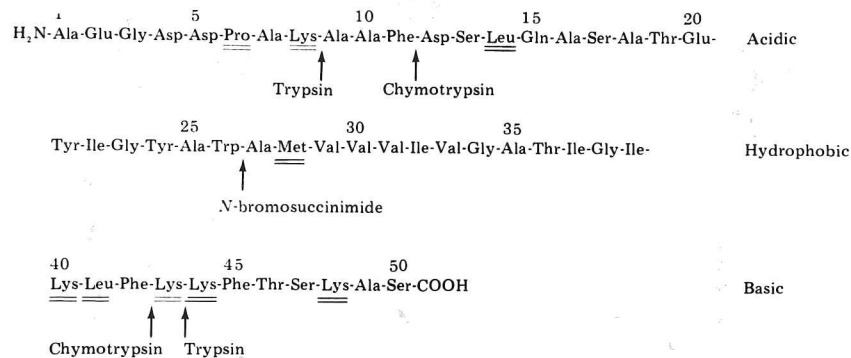


FIG. 1. Amino acid sequence of the M13 coat protein (7, 8). Amino acids which were labeled in the current study are underlined. Positions of cleavage by trypsin, chymotrypsin, and N-bromosuccinimide are indicated by arrows.

Figure 1.7. The amino-acid sequence of the g8p of bacteriophage fd. Three distinct zones are visible: an acidic zone from the N-terminus; a central hydrophobic zone thought to insert into the membranes of infected cells; and a basic zone towards the C-terminus, which lines the capsid and may interact with the DNA. The cleavage sites of some enzymes and a reagent are shown.

(From Wickner, 1976.)



disassembly has been reported with 2-chloroethanol (Ikehara & Utiyama, 1975), but the work has not proven repeatable (Dr. J. Armstrong, pers. comm.). The most successful work has shown tryptic protection of g8p incorporated into phospholipid vesicles from detergent solutions (Wickner, 1976; Wickner, 1970). The protein was inserted in the same orientation as in vivo with the same membrane spanning region (Wickner, 1975; Woolford & Webster, 1975). The CD spectra of these preparations showed the g8p conformation was no longer 100% alpha-helix but rather 50% alpha-helix, 30% beta-pleat and 20% random coil (Chamberlain et al., 1978; Williams & Dunker, 1977; Zwizinski & Wickner, 1977; Makino et al., 1975). This is in good agreement with the structure predicted from the amino-acid sequence, and also adopted in solutions of dilute SDS or sodium deoxycholate (Nozaki et al., 1978; Nozaki et al., 1976).

A potentially similar structural transition has been noted at a water/chloroform interface. The capsid proteins rearrange into a hollow spheroid of 39 nm diameter, exposing the DNA from a circular aperture (Griffith et al., 1981). An intermediate morphology has been isolated at 2°C, the I-particle, which is swollen not spherical. Re-exposure to the hydrophobic interface at room temperature completes the transition, which appears to be a concerted conformational change. Whether this is more akin to disassembly or denaturation is not known (Lopez & Webster, 1982; Manning et al., 1981).



## 1.5.3 DNA Replication

An initial problem in the replication of viral nucleic acid, once it has reached the cytoplasm, is that the DNA is single-stranded whereas the host cell enzymes deal with double-stranded DNA. Thus the viral (+) strand must be converted to a duplex replicative form (RF), from which progeny (+) strands may be synthesised for encapsidation later in infection. It may be enough to account for this bizarre sequence as an accident of nature, but the consequent halving of the energy and materials demand in replication may well be a significant evolutionary advantage.

The process generating RF duplexes is inhibited by rifampicin, indicating that RNA polymerase activity is required; strains with rifampicin-resistant RNA polymerases do not exhibit this inhibition (Brutlag *et al.*, 1971). The single-stranded DNA may be complexed with *E.coli* single-strand binding protein (SSB, DBP or HDP) *in vitro* to give a pre-initiation complex, with a beaded appearance in electron micrographs (Sigal *et al.*, 1972). A period of RNA synthesis is then necessary before the DNA is competent to replicate (Schaller, 1978; Wickner, 1972). The primer synthesized is 30 nucleotides long starting with pppA at position 5739, which is in the intergenic zone on hairpin E (Geider *et al.*, 1978). Thus SSB binding is thought to leave the hairpin free for polymerase binding (Geider & Kornberg, 1974). In support of this, RNA polymerase binding is found to protect the region 5610 to 5740 bases from nuclease digestion, and the electrophoretic mobility of this

protected fragment suggests it is base-paired (Gray et al., 1978; Niyogu & Mitra, 1978; Schaller, 1978). The SSB sharpens primer termination, for naked DNA gives a primer with a ragged end (Geider et al., 1978).

The host cell replisome then binds at the primer and elongates the (-) DNA strand from there, eventually coming full circle to replace it with DNA (McHenry & Kornberg, 1977). Ligase seals the strands together to give a relaxed, double-stranded circular DNA of viral (+) and complementary (-) strands (Fig. 1.8). This is RF IV, which is converted to the supertwisted RF I by host cell DNA gyrase, as shown by the passage of pulse-labelled DNA from one form to the other unless the enzyme is inhibited with novobiocin (Horiuchi et al., 1978; Horiuchi & Zinder, 1976). Thus RF IV grades into RF I as superhelical turns are added, a spectrum of topoisomers being formed as intermediates.

The superhelical RF I is essential to viral replication in that it is the template rendered active by the protein product of gene 2 (Horiuchi et al., 1978). This is a site-specific nicking enzyme, which initiates rounds of replication by generating the nicked (and thus relaxed) duplex of RF II (Lin & Pratt, 1972). It has been isolated as a globular protein of  $M_r$  45,000 (Meyer & Geider, 1979a). The nick can only be made at the correct site in a supertwisted duplex DNA, which effectively reserves the nicking activity to RF I (Schaller, 1978; Fidanián & Ray, 1972). In vitro the recognition sequence was found to be 5'CTTT/AAAT3' (cut at /) leaving a 5'P and 3'OH at positions 5781/5782 on the (+) strand of the DNA (Harth et al., 1981; Meyer et al., 1979; Meyer & Geider, 1979b;

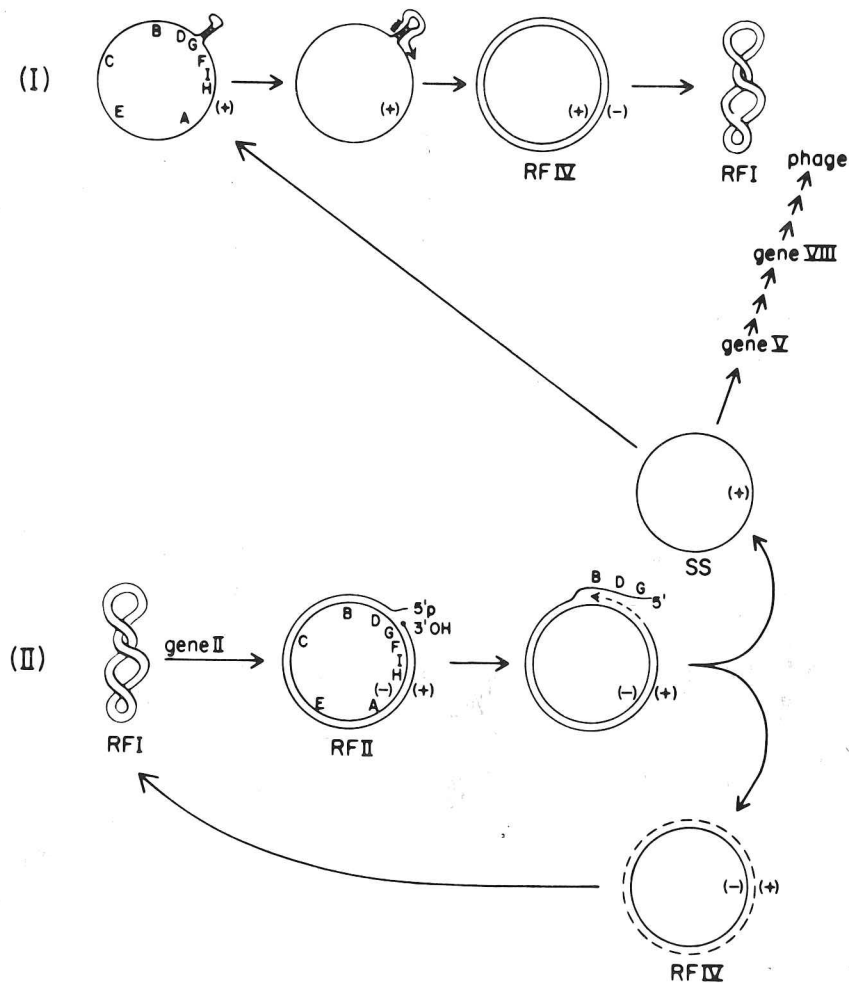


Figure 1.8. The pattern of DNA replication in filamentous bacteriophages. Letters on the circles represent the positions of the Hae III fragments. The initial priming on hairpin E is followed by DNA elongation on the (+) strand template to give a relaxed double-stranded DNA (RF IV), which is converted to the supercoiled RF I by DNA gyrase. This replicates to give one single-strand and one RF IV, both of which may then re-enter the cycle. Late in the course of infection, the (+) strands are sequestered by g5p before assembly into phages.

(From Schaller, 1978.)

Geider & Meyer, 1978). This is in the intergenic zone, at the end of hairpin D, but the secondary structure does not appear to be necessary for the nicking activity (Dotto et al., 1982). Full activity requires 12 nucleotides from the 5' end of the site to be intact, and the presence of a second regulatory zone 89-125 nucleotides from the nick (Cleary & Ray, 1981).

A second activity has been noted in this phage-coded enzyme. At high concentrations of magnesium ions it can nick, relax and religate the supercoiled RF I to the relaxed RF IV, though RF II (nicked and relaxed) is not a substrate for the religation. Consequently, overproducing plasmids in gene 2 confer Ff-resistance, because the RF I is relaxed faster than it can be supertwisted, so there is no substrate to be activated for replication (Dotto et al., 1981a).

The activated (nicked) RF II possesses a free 3' end of nucleic acid base-paired with the (-) strand. This is effectively a primer, the (+) strand origin of replication, from which the host replisome proceeds to replicate DNA (Dotto et al., 1981b; Meyer & Geider, 1981). The old (+) strand is displaced as a new (+) strand is polymerized in its place, so the process is a rolling circle replication. The displaced strand is dependent on g2p for its correct cleavage into haploid genome lengths and simultaneous ligation. Failure to cleave correctly gives rise to virions of multiple lengths containing a single, multiple-length loop of DNA. These viruses cannot replicate stably, for the origin and terminus of replication must be coincident (Dotto et al., 1984); otherwise, any

DNA between them would be lost in the first round of replication (Dotto & Horiuchi, 1981; Cleary & Ray, 1980; Horiuchi, 1980; Horiuchi & Ray, 1979).

The (-) origin of replication, on the other hand, is displaced with respect to the (+) origin. It is only exposed for activity once the (+) strand is almost completed, so the products of the rolling circle step are one (+) strand and one relaxed, ligated duplex - an RF IV. The RF IV can then proceed through the cycle again, being supertwisted to RF I, nicked by g2p to RF II which replicates to return a RF IV and a (+) strand. Thus the net product of each round of replication is one (+) strand (Schaller, 1978). This is complexed by SSB to give the pre-initiation complex, following the pathway outlined above to give one RF IV. This somewhat tortuous route allows one RF IV to give rise, by asymmetrical replication, to two RF IV and thus causes exponential multiplication of replicative forms (Fig. 1.8). In log-phase cells a ratio of 2:1:7 RFI/RFII/(+) strands is found, suggesting that the rolling circle replication is much faster than the RF interconversions or duplexing of (+) strands (Marvin & Hohn, 1969).

#### 1.5.4 Protein Synthesis

Protein synthesis must proceed in tandem with DNA replication, for the amount of protein available for the assembly of virions must match the available quantity of DNA - otherwise one or other component will accumulate rapidly inside the host cell. Complete inhibition of host

cell protein synthesis is not observed - and is undesirable as a healthy host will continue to be a suitable environment for viral multiplication.

Hybridization experiments show that the RNA formed falls into a graded series of sizes, the longer sequences containing the shorter ones. It was suggested that these arose from several independent promoters around the genome, synthesizing overlapping mRNAs which all terminated at a common point (Hulsebos & Schoenmakers, 1978; Chan et al., 1975). In support of this eight promoters have been identified together with a single rho-independent terminator (Moses & Model, 1984; Edens et al., 1978; Sugimoto et al., 1977). Thus genes present a long way from the terminator would be present in fewer transcripts - and hence be less well expressed - than those closer to it. This is known as the cascade model of transcriptional control. A modulation in the pattern of expression is caused by the differing half-lives of the transcripts (le Farina & Model, 1978).

A peculiarity is found in gene 3, which resides just distal to the terminator but lacks a promoter. Its expression seems to depend on non-termination. Gene 7 has also posed a problem for the cascade theory, as it is expressed poorly but lies between two well expressed genes. It is thought that although a commensurate amount of mRNA is available, the poor ribosomal binding site ahead of the gene is responsible for its lack of expression (Hulsebos & Schoenmakers, 1978). The expression of gene 8 is not limited by transcriptional cascading as initiation at its

own promoter gives full expression, even if this gene is cloned into the intergenic zone (Moses & Horiuchi, 1982; Moses et al., 1980; Edens et al., 1978).

The sequence of the gene 8 transcript has been determined (Sugimoto et al., 1977). It shows a 23 residue amino-terminal extension on the protein which is cleaved soon after membrane insertion. The sequence is thus thought to represent a signal sequence which directs insertion into membranes (Blobel & Doberstein, 1978). In vitro synthesis gave the water soluble pro-coat protein (g8p), which was reported to be capable of subsequent insertion into pure phospholipid vesicles (Wickner et al., 1978). Other experiments, however, indicated that a delay between protein synthesis and vesicle addition prevented membrane insertion, suggesting that insertion was co-translational (Chang et al., 1979). Studies in vivo have followed the membrane insertion of radiolabelled protein synthesised on free polysomes (Date et al., 1980a; Smilowitz et al., 1972). Addition of chloramphenicol - stopping protein synthesis - does not inhibit the accumulation of pro-coat protein in the membrane (Ito et al., 1980). The insertion thus appears to be post-translational, though the competence of the protein to insert diminishes with time (Goodman et al., 1981; Watts et al., 1981). Comparison of sequence data from the gene and the mature protein indicates that the g3p of Ff phages also has an N-terminal extension which is cleaved on membrane insertion. No post-translational modification of other phage proteins has been described.



Membrane insertion is accelerated by an electric field across the membrane. The process is inhibited by uncouplers (DNP, CCCP) leading to the accumulation of pro-coat protein on the inner face of the plasma membrane (Wickner, 1983; Date et al., 1980a,b). The electric field could help insertion, as the acidic (negatively charged) portion of g8p is the one which traverses the membrane, passing down its electrochemical gradient (Wickner, 1983).

Once inserted the protein appears to be surrounded by a lipid annulus - primarily of cardiolipin (Silver et al., 1981; Mandel & Wickner, 1979; Chamberlain & Webster, 1976). Processing to the mature form may well be a limiting factor in g8p biosynthesis as plasmids enhancing signal peptidase activity accelerate the accumulation of g8p in the membranes of infected cells (Date & Wickner, 1981). The g8p is, in fact, toxic to the cell for accumulation of g8p in the plasma membrane leads to the formation of great multi-lamellar whorls of membrane in the host cell, which eventually dies. This toxic effect is thought to be due to interference between the g8p and normal membrane proteins (Schwartz & Zinder, 1967).

#### 1.5.5 The Role of Gene 5 Protein

Gene 5 protein is a globular cytoplasmic protein found in cells infected with filamentous phage about 10-11 minutes after penetration (Mazur & Zinder, 1978; Nakashima et al., 1978a,b). Its overall effect is to bind single-stranded DNA and thus to inhibit RF formation, favouring



the net synthesis of (+) strands. These (+) strands are complexed with g5p as a filamentous aggregate visible in electron micrographs of gently lysed cells. Such strands are subsequently assembled into mature virions, so the g5p.DNA nucleoprotein may be regarded as the pre-assembly complex. The g5p competes with SSB for the single-stranded product of each round of replication. Early in infection the SSB is in excess, so a pre-initiation complex is formed and the (+) strand is channelled back into RF replication, giving the exponential multiplication described previously. Later the level of g5p rises and removes (+) strands from the replicative cycle, the latter now acting mainly as a source of viral strands. Replication has two distinct phases, one involving the multiplication of RF, the other generating (+) strands for incorporation into progeny phages (Lerner & Model, 1981; Horiuchi et al., 1978; Geider & Kornberg, 1974; Pratt & Erdahl, 1968).

The complex isolated from infected cells is a helical nucleoprotein aggregate, about 1000 nm long (Fig. 1.2 & Fig.1.9), having about 5 nucleotides per subunit of g5p (Pratt et al., 1974). In vitro reassociation of the complex gives rise to a nucleoprotein with 4 nucleotides per subunit (Torbet et al., 1981; Anderson et al., 1975; Pretorius et al., 1975). The complexes also show morphological differences: in vivo the complex is linear, whereas in vitro it is stellate (Zentgraf et al., 1977). This can be explained if initiation of assembly in vitro occurs at several sites on the DNA, whereas in vivo the assembly propagates from a single site. The specificity of the system in vivo is presumably conferred by some extrinsic agency such as

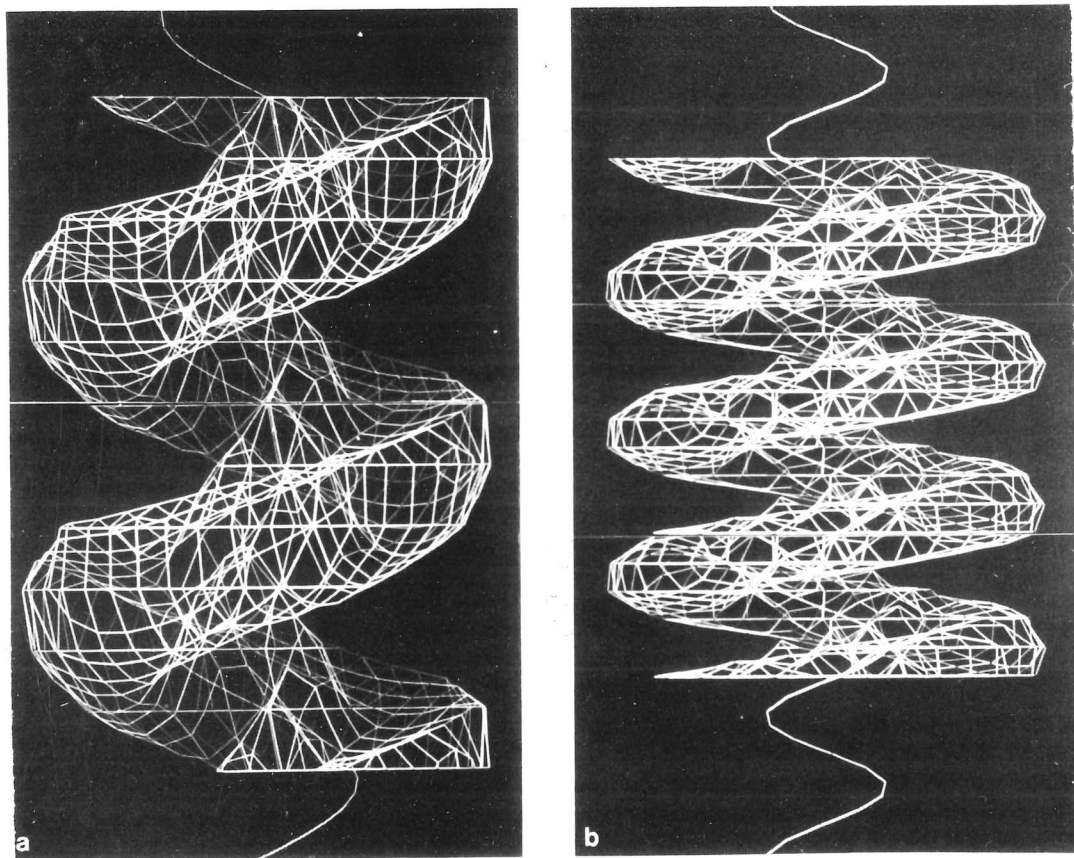


Figure 1.9. The helical structures of the g5p.DNA complexes of (a) fd and (b) Pf1. These computer projections were based on high resolution electron micrographs coupled with image processing by computer. The fd complex has a larger pitch and a more open form (see also Fig. 1.2).

(From Gray et al., 1982.)

the directionality of DNA synthesis. The structural subunit of the complex is the g5p dimer, which binds to the DNA co-operatively (Alma et al., 1983; Cavalieri et al., 1976). This co-operativity is negative: that is, lone dimers fail to bind DNA stably but those binding at the elongating end of the complex are retained (McGhee & von Hippel, 1974; Chang & Weiskopf, 1973). Precisely how these dimers associate with the DNA and each other to form complexes is not known.

Both the DNA sequence and the protein sequence of the g5p are known. Reported attempts to predict the secondary structure have so far been imprecise (Anderson et al., 1975). X-ray diffraction studies of the crystallised Ff protein have shown it to be a globular protein with only beta-pleated structure. A groove 3 nm long under a twisted three-ply beta-sheet was suggested as the DNA binding site, while one beta-ribbon appears to mediate dimerisation and the other promotes inter-subunit contacts in the complex (McPherson et al., 1979).

There has been considerable interest in the literature about the specific groups responsible for binding DNA in such proteins (Dimcoli & Helene, 1974; Chang & Weiskopf, 1973; Gabbay et al., 1973; Helene et al., 1973). Two types of interaction have been described: the electrostatic interactions between positively charged amino-acid residues (principally lysine) and the phosphate backbone of nucleic acids, of their nature non-specific; and the interaction of non-polar residues such as tyrosine and tryptophan with the bases of DNA, which may allow sequence specificity.

In the case of g5p.DNA complex, Tyr-21, -41, -56 are protected from the solvent and from nitration reactions when in the complex (Day, 1973). These residues are exposed in the putative DNA binding groove shown by X-ray crystallography, as are Phe-13 and Cys-33. The latter has been shown to be in close proximity with the DNA by UV-crosslinking (Paradiso & Konigsberg, 1982; Paradiso et al., 1979). The DNA may be bound primarily by the intercalation of Tyr-41, Tyr-56 and Phe-13 with DNA bases, but the close proximity of Ser-75 to the DNA, coupled with the salt-dissociability of the complex, implies that electrostatic interactions are also significant (Kneale & Tsugita, 1985; Kneale, 1983; McPherson et al., 1979).

The gross structure of the complex has been determined by high-resolution electron microscopy coupled with computer image processing techniques (Gray et al., 1982b). The X-ray data were interpreted to show that the DNA was wound around the outside of a g5p rod (McPherson et al., 1979), while the evidence from neutron scattering proved that some protein density occurred beyond the maximal radius of DNA density (Gray et al., 1982a). It has now been shown that the complexes of both phage classes are helical, with the DNA running inside the surface of the protein helix. The Ff complex has about 97 turns of helix in a length of 880 nm, while the Pf1 complex shows 185 turns in 964 nm (Gray et al., 1982a). The differences in symmetry between the g5p.DNA complexes reflect differences in the structures of the phages assembled from them.

## 1.5.6 Assembly

The g5p.DNA complex of filamentous phages is a means of removing (+) strands from the replicative cycle, to sequester them for incorporation into progeny phages. Since the complex is helical, it also has the effect of supertwisting the DNA to a degree similar to that found in the mature virion. Thus it prepares the DNA for assembly into virions.

Release of viral progeny is accomplished without cell lysis (Hofschneider & Preuss, 1963). There is some evidence that the outer membrane is damaged by infection, for some lipopolysaccharide is released to the medium (Roy & Mitra, 1970a,b; Falaschi & Kornberg, 1965). Phage release begins 10-20 minutes after infection, depending on the metabolic state of the host cell (Brown & Dowell, 1968; Pratt & Erdahl, 1968). Progeny can be seen in electron micrographs to be extruded from a few points on the cell surface, thought to be the perforations in the cell wall through which the pili normally protrude (Fig. 1.10).

During assembly, the viral nucleic acid must shed g5p, penetrate the membrane and pick up capsid protein from the membrane (Wickner & Killich, 1977). How or when the minor capsid components attach to the DNA or capsid is not yet known (Fig. 1.10). It has been proposed that the g8p in the membrane has lateral interactions similar to those in the capsid, and that as capsid is extruded from the membrane surface only a small rearrangement of subunit interactions is necessary (Marvin &

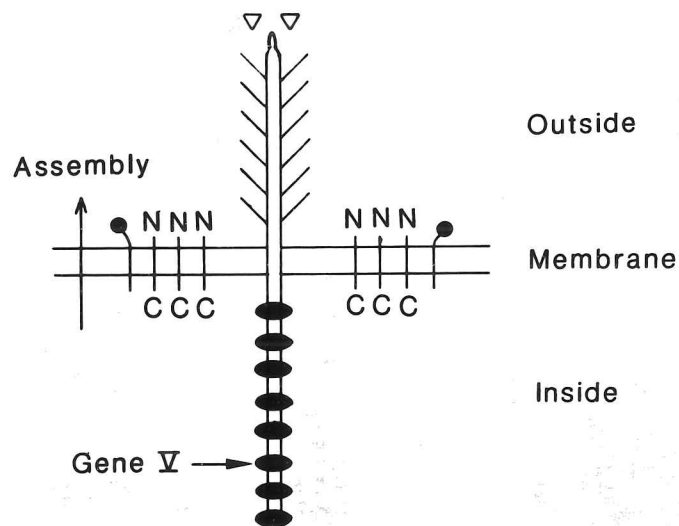




Figure 1.10. The assembly of bacteriophage fd at the bacterial membrane. As the DNA passes through the membrane it sheds g5p and is simultaneously encapsidated in g8p ( / ). The g3p, with its globular domain (  ), is shown awaiting assembly in the membrane. The g7p.g9p is shown at the end of the elongating capsid (  ), and the location of g6p during assembly is not indicated. It is not known at what stage these components become associated with the DNA.

Wachtel, 1976; Marvin & Wachtel, 1975). The transition of g8p from its membrane bound form to that in the capsid is the reverse of the change occurring in disassembly. In agreement with this is evidence that radiolabelled g8p from infecting phage which is dispersed into the membrane can subsequently be reassembled into progeny virions (Armstrong et al., 1983; Mandel & Wickner, 1979; Smilowitz, 1974).

In common with many such linear processes, assembly of filamentous phages may be subdivided into three phases: initiation, elongation and termination. Elongation, usually the longest and hence the easiest to study, has been described above. Attempts to recreate the process in vitro have been unsuccessful. Reaggregation of detergent-solubilised g8p with DNA produced a fibrous material with only 50% alpha-helix, conferring some resistance to nuclease digestion on the DNA (Knippers & Hoffman-Berling, 1960a,b). Similar material was prepared from g8p alone, showing that DNA protection had been non-specific and due simply to the trapping of DNA between the fibrous aggregates (Knippers & Hoffman-Berling, 1960c). Such aggregation is a parallel of the self-assembly of TMV and shows that much of the morphogenetic information resides in the coat protein.

Initiation of assembly must occur at the g7p.g9p-end of the virion, corresponding with the intergenic zone of the DNA, as it is the first region to emerge from the cell surface (Lopez & Webster, 1984; Armstrong et al., 1981). This region also seems to be the zone of attachment of the DNA to partially denatured virions (spheroids) suggesting the



possibility of DNA-protein binding at this site (Griffith et al., 1981). It has been speculated that the double-stranded hairpins, found at one end of the virion, may be involved in the initiation process (Ikoku & Hearst, 1981; Webster et al., 1981). Therefore the earlier model of assembly, with phage being extruded g3p end first (Kornberg, 1974), has been superseded. The g3p had been thought to be the initiator of morphogenesis, primarily because it was the only minor protein component known and thus imposed a gross asymmetry on the capsid.

### 1.6 Perspectives

As the preceeding attempts to show, the later stages of the life-cycles of filamentous bacteriophages are poorly understood. This is particularly surprising because of their simple construction compared with viruses such as the T-phages. Yet the complex infective and assembly processes of the T-phages are understood in some detail. A primary problem in the study of filamentous viral assembly is that it does not take place in the cytoplasm with many virions being assembled in parallel. The viral filament is only assembled at the last moment, at the membrane as the DNA is passing through it to the medium outside (Wickner, 1977). Such an oriented system of three phases (cytoplasm, membrane, medium) is difficult to arrange in vitro.

A further factor hampering development in this area is the lack of comparative information between the Class I and Class II phages. It is often very highly instructive to compare systems, whether from a



biological or physicochemical point of view, if they are functionally related. Similarities and differences can be related to common or contrasting structures and functions. For example, the different symmetries of the Ff and Pf1 capsids must relate to differences in the mechanics of assembly.

The recent discovery of minor protein components, in addition to g3p, in Ff capsids raises additional questions. Are corresponding proteins also present in the capsids of the Class II phages or is this a feature of difference between the two classes - perhaps related to the differences in symmetry? In this context it is notable that Class I phages show a five fold symmetry in the capsid which is continued into the minor proteins at the capsid ends. If there are minor proteins in the one-start Pf1 capsids, what restricts their stoichiometry? The analysis of the potential secondary structures of these proteins may also yield interesting information.

The presence of these minor proteins in the capsids of Ff phages merits their consideration as possible pilot proteins. If any of these control assembly, they will presumably be present on the g5p.DNA complexes. It would be of interest to isolate these complexes, and to look for the presence of minor proteins regulating the initiation of assembly. Furthermore, the phages assembled from these complexes contain DNA in a specific linear orientation, with the intergenic zone at one end. It would not be unreasonable to ask whether the g5p.DNA complexes contain DNA oriented in a similar manner. This thesis describes work aimed at answering these important questions about filamentous phage assembly and the underlying DNA-protein interactions.

## CHAPTER 2

### MATERIALS AND METHODS

## 2.1 Materials

### 2.1.1 Chemical Reagents

All chemicals used were of Analytical Reagent grade, obtained from Fisons, except:

From BDH:

Ammonium carbonate

Ammonium chloride

Bromophenol blue

Citric acid

Ethylene diamine tetra-acetic acid (GPR)

Formaldehyde

Glutaraldehyde solution (50%)

Hydrochloric acid

8-Hydroxyquinoline

Magnesium chloride

Magnesium sulphate

2-Mercaptoethanol

Potassium acetate

Sodium azide

(tri)-Sodium citrate

Sodium dodecyl sulphate

Sodium hydroxide

Sodium sulphate

	Sucrose
	TEMED (Lab. grade)
Sigma:	All (L)-amino-acids (99% pure)
	Ficoll 400
	Lysozyme (egg white), grade IV
	Metrizamide, grade I
	Phenyl-methyl-sulphonyl-fluoride
	Subtilisin type VIII
	Thiamine hydrochloride
	Trizma base, reagent grade
May & Baker:	Acetic acid (glacial,LR)
	Butan-2-ol (LR)
	Methanol (LR)
	di-Potassium hydrogen orthophosphate (AR)
Difco:	Bacto-agar
	Bacto-tryptone
	Yeast extract
Beckman:	555 rotor cleaning fluid
Boots:	Glucose
BRL:	Caesium chloride (optical grade)
ICI:	Chlorox disinfectant

J.Burroughs: Ethyl alcohol (AR)  
Koch Lite Ltd: 2,5-di-Phenyloxazole (puriss.)  
Lip Services: Lipsol  
Park Sci. Ltd: di-Thiothreitol

Radiolabelled amino-acids were obtained from Amersham International PLC:

L-[5(n)-<sup>3</sup>H] Arginine monohydrochloride  
1 mCi in 1 ml water, sterile.

L-[G-<sup>3</sup>H] Tryptophan  
1 mCi in 1 ml ethanol/water (1:1, v/v).

These were supplied as better than 95% pure and stored as recommended.

#### 2.1.2 Buffers and Media

The following buffers were frequently used:

NET: 100 mM sodium chloride, 10 mM Tris.HCl, 1 mM EDTA, pH 7.5  
TE: 10 mM Tris.HCl, 1 mM EDTA, pH 8.1  
NT: 100 mM sodium chloride, 10 mM Tris.HCl, pH 7.8

All solutions were brought to the correct pH, as necessary, using a Russell Tris-compatible electrode (TR-CMAWL-4-tb) from Russell pH Ltd., Auchtermuchty, Fife, Scotland. NET-buffer was used mainly for suspending phages, TE-buffer for g5p.DNA complexes and NT-buffer for cells.

Luria agar was used for maintaining cell lines in stabs and on plates. This consists of 1% Bacto-tryptone, 0.5% yeast extract, 1% sodium chloride, 0.1% glucose, 10% agar in glass distilled water and was prepared by autoclaving the components in suspension (Luria & Burrous, 1957). The rich medium 2xTY was used for growing cells in liquid culture. The composition was: 1.6% Bacto-tryptone, 1% yeast extract, 0.5% sodium chloride.

A defined medium, MTPA, was extensively used in the course of this work to provide a stringently controlled environment for radiolabelling cultures (Vinuela et al., 1967). The components of this medium were:

MTPA salts:	0.2 M	Tris.HCl
	19 mM	Sodium chloride
	200 mM	Potassium chloride
	40 mM	Ammonium chloride
	680 mM	Potassium di-hydrogen orthophosphate
	320 mM	Sodium sulphate
	5 mM	Calcium chloride
	2 mM	Tyrosine

Amino acids: Asp, Glu, Trp as 0.0333 M solutions.

The remaining 16 common amino-acids as 0.1 M solutions.

Glucose: 20%

Thiamine: as the hydrochloride, 10 mg/ml.

All components were sterilised by autoclaving except glutamine and thiamine, which were filter sterilised through a Millipore Millex-GV 0.22  $\mu$ m filter into autoclaved vessels. These components were combined to give a culture medium containing, per ml:

<u>Component</u>	<u>Volume (microlitres)</u>
MTPA salts.....	500
16 (0.1 M) amino-acids.....	10 each
3 (0.033 M) amino-acids.....	30 each
Glucose.....	18
Thiamine.....	1
Glass distilled water.....	232

### 2.1.3 Organisms

Stocks of bacteriophages fd and Pf1 were kind gifts from Dr. D.A. Marvin. Stock fd 44 was a suspension of 6 mg/ml phage fd in 100 mM Tris.HCl, pH 7.5, 0.1 mM PMSF and 0.02% sodium azide. Stock Pf1 52 was a suspension of 16 mg/ml phage Pf1 in 10 mM Tris.HCl, pH 8.1, 1 mM EDTA, 0.1 mM PMSF, 0.02% sodium azide. These were stored at 4°C.

Escherichia coli K38 was a gift from Dr. D. Hill. E.coli strains P4X SB167, MA 1013 and AT 12-55 (Genetic Stock Center numbers 5462, 5421 and 5607 respectively) were kindly provided (by Dr. B. Bachman of the E.coli Genetic Stock Center, Department of Human Genetics, Yale University, USA) as a few drops of glycerolised culture on filter paper discs. These were placed on Luria agar plates and streaked out with MTPA medium as diluent. All were HFr strains (thus pilus bearing), their auxotrophic requirements being:

P4X SB167.....Arg...Met

MA 1013.....Arg...Thiamine

AT 12-55.....Arg...Thr...Leu...Thiamine

Thus all were able to grow on the culture media used.

Pseudomonas aeruginosum K was a kind gift from Dr. D.A. Marvin.

Stocks were maintained on Luria agar plates. Longer term storage was provided by preparing stabs in Luria agar, kept at room temperature, and by storing cell pastes, pelleted in 50% glycerol at -20°C (Maniatis et al., 1982).

## 2.2 Plaque Assays

Plaque assays were performed by the double layer agar technique,



plaques showing up as clear spots of retarded growth on a greyish lawn of bacterial cells. The support agar contained, per litre:

Ammonium chloride.....	1.0 g
Glucose.....	4.0 g
Magnesium sulphate heptahydrate.....	0.2 g
di-Potassium hydrogen sulphate.....	12.2 g
tri-Sodium citrate.....	2.0 g
Sodium dihydrogen phosphate monohydrate.....	5.0 g
Difco-agar.....	15.0 g

This was autoclaved and used at the rate of about 25 ml per Petri-dish, bubbles being flamed away carefully to avoid confusion with plaques later. Dishes were prepared at least one week in advance to allow time for the condensation inside to evaporate.

The overlay agar was of similar composition but with only 3.6 g/l of Difco-agar and 90 ml/l of host cell culture in 2xTY at  $A_{680}$  0.3 (E.coli K38 for fd, P.aeruginosum K for Pf1). Aliquots of 3 ml were vortexed briefly with 0.2 ml samples then poured onto support agar and given 30 sec of lateral agitation. After being given around 30 min to set, plates were incubated overnight at 37°C. Samples were assayed in tenfold serial dilutions, plates with 25-250 plaques being counted. If quantification of phage particles was required, as in determinations of specific radioactivity, a plaquing efficiency of 80% was assumed on the basis of control assays done on the phage stocks (section 3.1).

## 2.3 Microbiology

### 2.3.1 Selective Radiolabelling of Cell Cultures

A colony of host cells was picked from a Luria plate, and inoculated into 2 ml MTPA medium. After incubation at 37°C overnight, with orbital shaking, the  $A_{680}$  of the culture was determined. MTPA (lacking the amino-acid to be radiolabelled) was prepared as 10 ml in a 40 ml culture tube, pre-warmed and aerated by shaking. The MTPA was inoculated with a volume of the overnight culture sufficient to give a starting  $A_{680}$  0.02. Culturing proceeded at 37°C with orbital shaking. The tube was examined periodically, samples being withdrawn as necessary to check the turbidity of the culture. At  $A_{680}$  0.16 the appropriate phages were introduced at a multiplicity of infection (m.o.i) 100, about  $10^{10}$  phages per ml of medium. At the same time, 1 mCi of the radiolabelled amino-acid, in sterile aqueous solution, was also added. Propagation continued for the requisite length of time (Lin et al., 1980).

### 2.3.2 Purification of Phages

Cells were pelleted from the culture by centrifugation in a SS34 rotor at 10,000 x g for 10 min at 4°C. The supernatant was decanted and brought to 5% polyethylene glycol (PEG) and 0.5 M sodium chloride by the addition of the finely ground solids. The suspension was allowed to stand overnight at 4°C, then the precipitate was collected by centrifugation at 10,000 x g for 90 min at 4°C. The pellet was

resuspended in 0.1% Sarkosyl in TE-buffer, and allowed to stand for 1 h to dissolve membranous debris. The precipitation step was then repeated, and the product resuspended in 5 ml NET-buffer. Phages were sedimented by centrifugation in an SW 50.1 rotor at 170,000 x g, for 5 h at 20°C. The resulting pellet of phages was resuspended in 0.1 ml NET-buffer by stirring the pellet with a small PTFE-coated magnetic spin-bar for 2 h (Yamamoto et al., 1970).

### 2.3.3 Purification of g5p.DNA Complexes

Cells were pelleted from the culture by centrifugation in a SS34 rotor at 10,000 rpm for 10 min at 4°C. The cell pellet was resuspended in 10 ml ice-cold NT-buffer by refluxing in a pasteur pipette. The cells were repelleted, resuspended, and repelleted to wash them. The pellet was resuspended in 0.25 of the original culture volume of TE-buffer. The  $A_{680}$  of this suspension was determined and the sample diluted to  $A_{680}$  1.0. To this was added 1/9th volume of freshly prepared 11 mg/ml lysozyme and 1/100th volume 10 mM PMSF (in ethanol). The mixture was stirred for 1 h at room temperature to accomplish cell lysis. Cells and debris were pelleted by centrifugation at 25,000 x g for 30 min at 4°C, and the complex bearing supernatant carefully decanted. The complex could then be collected by sedimentation at 100,000 x g for 3 h at 20°C.

## 2.4 Analysis of Proteins

### 2.4.1 Polyacrylamide Gel Electrophoresis

Analysis of proteins was performed by polyacrylamide gel electrophoresis in vertical gel-tanks after the method of Simons et al. (1979). Resolving gels were 1.2 mm thick with 15 cm track lengths. They contained 15% acrylamide, 0.4% bisacrylamide, 8 M urea, 0.1% SDS, 0.5 M Tris.HCl, pH 8.9. Stacking gels were cast with 5% acrylamide, 0.23% bisacrylamide, 8 M urea, 0.1% SDS, 0.5 M Tris.HCl, pH 8.9. The combs were siliconised to aid extraction from the adhesive but brittle gels. This was done by immersing them in dimethylchlorosilane solution for 5 min then washing in plenty of water.

Before casting, the gel solution was filtered to remove fluff and insoluble material. Polymerisation was effected with 0.0005 volumes TEMED and 0.005 volumes of 10% ammonium persulphate solution. The well buffer contained 6 M urea, 0.1% SDS, 77 mM glycine, 10 mM Tris, pH 8.9. It was always prepared fresh before use, as the pH was found to change on standing for a few days, giving diminished resolution.

Samples to be loaded were mixed with at least 5 volumes of cracking buffer (total 0.1 ml): 8 M urea, 2% SDS, 5% glycerol, 10% 2-mercaptoethanol, 62 mM Tris.HCl, pH 6.8, 0.01% bromophenol blue - then boiled in a water bath for 10 min to allow disaggregation and denaturation of the resilient phage proteins. Screw capped reaction

vials were used to prevent the escape of hydrogen sulphide into the laboratory, and to minimise water loss. In the event of a bad seal, urea would precipitate from its supersaturated solution as the sample cooled. This could be redissolved by topping up the sample with 10% glycerol and rewarming.

Electrophoresis was carried out overnight at a constant voltage sufficient to give a current of 12 mA at the outset. This dropped to around 7 mA by the morning, causing bands to move more slowly and reducing the chances of an over-run gel. Better resolution was obtained by such slow electrophoresis than with faster runs at higher voltages and currents.

#### 2.4.2 Calibration Proteins

These were run to calibrate the gel system for relative molecular mass (Fig. 2.1). The composition is shown below, made up from stock protein solutions at 0.7 mg/ml in 0.0625 M Tris.HCl, pH 6.8, 10% glycerol, 5% 2-mercaptoethanol, 2% SDS (courtesy of Dr. L. Packman, Dept. of Biochemistry, Cambridge).

<u>Component</u>	<u>Volume (microlitres)</u>	<u>M<sub>r</sub></u>
Bovine Serum Albumin.....	30.....	66,000
Ovalbumin.....	60.....	45,000
Carbonic anhydrase.....	30.....	31,000



Figure 2.1. The pattern of electrophoresis of the calibration standards used on polyacrylamide gels. From top to bottom the bands are: Bovine Serum Albumin, Ovalbumin, Carbonic Anhydrase, Myoglobin (faint), Lysozyme, Cytochrome c, Aprotinin, Insulin heavy and light chains.

Myoglobin.....	30.....	17,400
Lysozyme.....	30.....	14,300
Cytochrome c.....	30.....	12,500
Aprotinin.....	10.....	6,500
Insulin.....	80.....	3,400
		2,340

The variation of concentration helps balance the intensities with which silver-staining shows the bands on a gel (Fig. 2.1).

#### 2.4.3 Silver-staining

A method based on that of Morrissey *et al.* (1981) was used; staining was performed in clean glass dishes. At no stage would a gel be touched, even with the gloved hand, as prints would develop later. Incubations of 30 min were performed in each of the following solutions:

1. 50% methanol, 10% ethanoic acid in water,
2. 7% methanol, 5% ethanoic acid in water,
3. 10% glutaraldehyde,
4. 4 rinses of water,
5. 5 mg/l dithiothreitol,
6. 0.1% silver nitrate,
7. 8.16 g sodium bicarbonate, 0.125 ml formaldehyde in 250 ml water.

Development was stopped with 8.4 g citric acid in about 25 ml water. Agitation was found to be particularly important in step 2, as the gels

used tended to float on the surface of the solution and show patchy staining as a result. Background staining can be virtually eliminated by thorough rinsing with two batches of 200 ml glass distilled water before development (step 7).

Fixing of gels as originally recommended was found to be poor. Neutralisation of the solution was found to be more effective, using 8.4 g of citric acid. Gels fixed in this way could be stored for years in sealed polythene bags, whereas those fixed in less citric acid darkened slowly.

#### 2.4.4 Fluorography

Gels to be fluorographed were immersed twice, with gentle agitation, in 15-20 volumes of glacial ethanoic acid for 15 min each time to remove all traces of water. A three hour immersion in 20% 2,5 diphenyloxazole (PPO) followed (in glacial ethanoic acid), still with agitation. The solution was aspirated back into its storage container, and the dish flooded with ice-cold water to precipitate the PPO in the gel. The gel was allowed to soak in a slow stream of water overnight ( about 1 l/h) to remove all traces of ethanoic acid, then dried down onto a paper support using a freeze-drier. Warming the gel with a Phillips 250 W infra-red lamp at 50 cm speeded the process up to 3 h. The stock solution of PPO was replenished to 20% with an amount of solid PPO depending on the gel volume.



This method differs from that of Skinner & Griswold (1983) from which it is derived, in that their initial 5 min incubation was not found sufficient to remove all water from the highly crosslinked gels used, so a total immersion of 30 min was substituted. PPO precipitation was in iced water to ensure the smallest possible PPO crystals and thus the highest resolution and sensitivity. Subsequent aqueous extraction of ethanoic acid from the gel was also prolonged as this was found to be necessary.

Gels were exposed to preflashed Fuji RX safety film for one day to a week at  $-70^{\circ}\text{C}$  and developed with the standard Kodak process for the film. PPO was recycled by precipitation from ethanol (with water) giving virtually complete recovery of fine white crystals (Laskey & Mills, 1975).

## 2.5 Reagents for DNA

### DNase I buffer:

- 10 mM Tris.HCl, pH 7.6
- 20 mM magnesium chloride.

### Low salt restriction buffer:

- 10 mM Tris.HCl, pH 7.6
- 20 mM magnesium chloride
- 1 mM dithiothreitol (added just before use).

## Medium salt restriction buffer:

100 mM sodium chloride  
10 mM Tris.HCl, pH 7.6  
20 mM magnesium chloride  
1 mM dithiothreitol (added just before use).

## Sma I buffer:

150 mM Tris.HCl, pH 8.0  
50 mM magnesium chloride  
150 mM potassium chloride  
Dilute tenfold in use.

## 'Phenol':

Phenol was melted and 0.1% 8-hydroxyquinoline added. The yellow phase was extracted once with 0.1 M Tris.HCl, pH 8.0, the phases separated by centrifugation at 5,000 x g for 5 min, then extracted with 0.1 M Tris.HCl, pH 8.0, 0.2% 2-mercaptoethanol. The product was stored under the latter buffer for periods of up to one month, or at -20°C for longer periods.

## 'Chloroform':

Prepared as 24 parts chloroform to 1 part isoamyl alcohol by volume

## 'Phenol/chloroform':

Prepared as a 1:1 mixture (v/v) prior to use and centrifuged briefly in an Eppendorf centrifuge to

separate the small excess of water excluded from the 'phenol' by the 'chloroform'.

## 2.6 Deproteinisation of DNA

The above reagents were used to deproteinise samples of DNA. The sample was placed in an Eppendorf reaction vial, and an equal volume of 'phenol' added. The mixture was vortexed briefly, allowed to stand for 1 min, then centrifuged at 10,000 x g for 1 min to separate the phases. The upper, aqueous layer was transferred to another reaction vial, extracted with 'phenol/chloroform' in the same way, and the upper layer transferred again to allow extraction with 'chloroform'.

The upper phase from this was removed to another reaction vial, and extracted with two volumes of water-saturated ether. The etheric top phase was discarded and the extraction repeated to remove all traces of phenol and chloroform from the sample which might otherwise interfere with the subsequent processing of the DNA (Maniatis *et al.*, 1982).

## 2.7 Agarose Gel Electrophoresis

Electrophoresis of DNA was usually performed in gels of 1% agarose, in horizontal gel tanks. The gel volume was 100 ml, cast in the running buffer of 0.089 M Tris, 0.089 M boric acid, 2mM EDTA, 5 mg/l Ethidium bromide (Maniatis *et al.*, 1982). Samples were prepared for loading by chelation of magnesium ions with an appropriate amount of EDTA (in

twofold stoichiometric excess) and the addition of 1/5th volume of loading buffer (15% Ficoll 400, 0.25% bromophenol blue in water). The detailed methodology followed Maniatis et al. (1982).

Gels were run with 5 mg/l Ethidium bromide incorporated into both gel and running buffer. Visualization was on a Fotodyne 3-3002 transilluminator. Background fluorescence was removed by rinsing the gel in glass distilled water for 15 min. (Tap water contains sufficient magnesium to cause destaining of the bands). Photographs were taken with a Polaroid CU-5 88-48 camera with its hood/stand at f 4.7, 2-6 sec depending on the brightness of the gel. Polaroid type 665 pos/neg film was used and processed according to the instructions in the pack.

### 2.8 Dialysis

Dialysis tubing (Visking type) was obtained from Medicell International Ltd. This was boiled in a large volume of 20% sodium hydrogen carbonate, 1 mM EDTA for 10 min, then rinsed thoroughly in glass distilled water. It was then autoclaved for 10 min in the glass storage container full of glass distilled water. After this had cooled, sodium azide was added to 0.02% and the tubing stored at 4°C.

Dialysis was normally performed in 4 l of buffer, with gentle stirring, at 4°C. If the removal of caesium chloride or Metrizamide was needed, from a concentrated solution, the first two dialyses were performed at room temperature to prevent crystallisation of the solute. Four rounds were usually employed, the last being against glass distilled water if the sample was to be freeze-dried.

### 2.9 Scintillation Counting

Samples were counted in a scintillant of 5% PPO in 1:2 Triton/toluene. An LKB Rackbeta scintillation counter was used. Quench correction was by the external standard reference method, giving output in dpm.

## CHAPTER 3

### PREPARATION OF BACTERIOPHAGES

### 3.1 Basic Bacteriology and Virology

Filamentous bacteriophages consist essentially of DNA molecules enveloped by protein capsids. To investigate these proteins, the phages must be produced efficiently from cultures of infected cells. The more crude phage product the culture can provide the more steps can be used, if necessary, in the purification process. Each minor protein component of the capsid, as reported in the literature, constitutes only 0.2% of the viral protein (mol/mol) (Lin et al., 1980). Consequently, they can only be detected reliably in viral preparations of very high purity. The purification methodology and the gel systems for analysing the coat proteins were therefore first studied for the phages from Class I. It was hoped that a similar approach could then be used to investigate the potential presence of such proteins in the Class II phages and the intracellular precursors of both groups.

Bacterial strains were cultured in either of two media, according to requirements. The rich broth 2xTY was used for strain propagation and experiments in which high yields of cells or phages was the prime consideration. In experiments involving radiolabelling, the closely defined medium MTPA was used (Vinuela et al., 1967). Escherichia coli K38 was found to grow more rapidly than Pseudomonas aeruginosum K on both these media. E.coli grew as a white (MTPA) or buff (2xTY) suspension, whereas P.aeruginosum was flesh-pink when pelleted from 2xTY. When grown in MTPA P.aeruginosum produced a white cell pellet from a yellow-green supernatant, this characteristic colour appearing in stationary phase growth.

Cultures were initiated by inoculating sterile medium with a sterile nichrome wire loop, from strains streaked out on Luria agar plates. These plates were found to have a fairly short life - a couple of weeks at most for P.aeruginosum. Longer term storage was thus necessary as agar stabs - more successful for P.aeruginosum - and glycerolised cells at -20°C - more successful with E.coli.

Initially, fd and Pf1 phages were checked for their ability to infect their hosts. Phage stocks were assayed by spectrophotometry in the UV region to show 6 mg/ml of fd and 16 mg/ml of Pf1 in the stocks used throughout this work (Appendix I). Given the relative particle masses of the virions these values were converted to phages per ml, and compared with the number of infective particles per ml derived from plaque assays. The plaquing efficiencies obtained, about 80% for fd and 25% for Pf1, were reasonable (Dr. D.A. Marvin, pers. comm.). The low efficiency of Pf1 may be related to the greater susceptibility of the capsids to shearing damage, as they are twice as long as those of fd.

### 3.2 Phage Biosynthesis

The filamentous phages are unusual in that progeny virus is liberated without causing lysis of the host cell (Hoffman-Berling & Mazé, 1964; Hofschneider & Preuss, 1963). According to the literature available, this process does not continue ad infinitum - there is a specific period of phage release between 20 and 120 min after infection (Marvin & Hohn, 1969). Thereafter phage biosynthesis is minimal, as the cells have



entered stationary phase. Previous practice in this laboratory, however, is in disagreement with these findings since culture times of up to 9 h are recommended in existing protocols (Armstrong, 1981). In order to determine whether such culturing is superfluous the timecourse of phage release was followed in a culture. Of course, this experiment has more fundamental implications, for it would be extremely unusual to find a virus which could release progeny virions over an indefinite period.

A culture of E.coli in 100 ml MTPA medium was prepared, starting with an inoculum from an overnight culture. At  $A_{680}$  0.16 fd phages were added at multiplicity of infection (moi) 100,  $10^{12}$  phages total. This moi ensures rapid adsorption, so the cells are infected synchronously. Under sterile conditions, samples of 0.5 ml were withdrawn into Eppendorf reaction vials every hour thereafter. Of each sample, 0.1 ml was used to determine  $A_{680}$ . This was needed to ascertain the status of the culture, following the pattern of exponential (log-phase) growth then deceleration into stationary phase. The remaining 0.4 ml of each sample was cleared of cells by centrifugation at  $5,000 \times g$  for 10 mins and the supernatant reserved for plaque assay. Thus it was hoped to correlate phage biosynthesis with the growth state of the culture. The results are shown in Figure 3.1. Phage release is roughly linear with respect to time, and continues at an approximately constant rate even during several hours of stationary phase culturing.

Although they agree with previous practice in this laboratory (Dr. J. Armstrong, pers. comm.), these results are in direct conflict with those

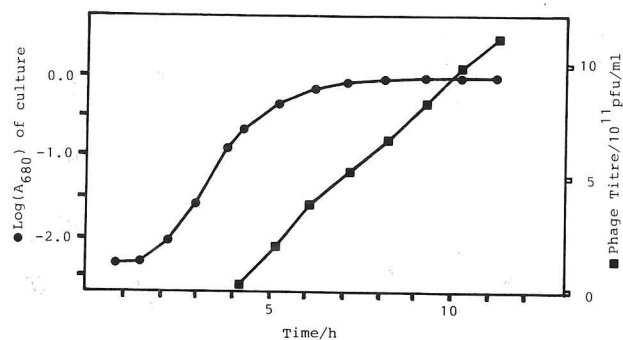


Figure 3.1. The biosynthesis of bacteriophage fd. A 100 ml culture of E.coli K38 in MTPA medium was infected with phage fd at moi 100. Samples were withdrawn at hourly intervals and the absorbance measured (circles) to follow the growth state of the culture. The quantity of virus present was also determined from these samples, by plaque assay (squares). Phage release is roughly linear with respect to time, and continues for several hours after the cells have reached stationary phase.

reported in the literature, both directly and by implication in the form of culture times in protocols (Roy & Mitra, 1970a; Marvin & Hohn, 1969). Two of the relevant graphs are re-presented (Fig. 3.2). Note the marked decline in the rate of phage release as the cells enter stationary phase. This was ascribed to a limitation in the availability of DNA replicating enzymes which comes about as cell division slows down. The course of infection was described as following a triphasic pattern: the initial infection, the rapid replication of RF's, and the release of virions. The final phase appears to give a burst of phage release after which the cells, in stationary phase, are no longer competent to synthesise phages (Dotto & Horiuchi, 1981; Marvin & Hohn, 1969; Pratt & Erdahl, 1968).

The resolution of this apparent dilemma hinges on the scales used in presenting the results. Re-examination of the results presented both in the literature and in this work shows that the re-presented graphs have a  $\log(\text{pfu})$  scale to be comparable with the  $\log(\text{cell density})$  scale - used to follow the progress of a culture through large increases in cell density. While cells grow exponentially, cell density gives rise to a straight line on the log scale. In stationary phase, this line is parallel to the ordinate, for cell density becomes roughly constant. The plot of  $\log(\text{pfu})$  does not: it continues to rise slowly, in a form reminiscent of a logarithmic curve. If the apparent decline in phage biosynthesis is really due only to the axes used, replotting the graphs found in the literature on linear axes should reveal linear relationships. The results of this replotting are presented (Fig. 3.3).

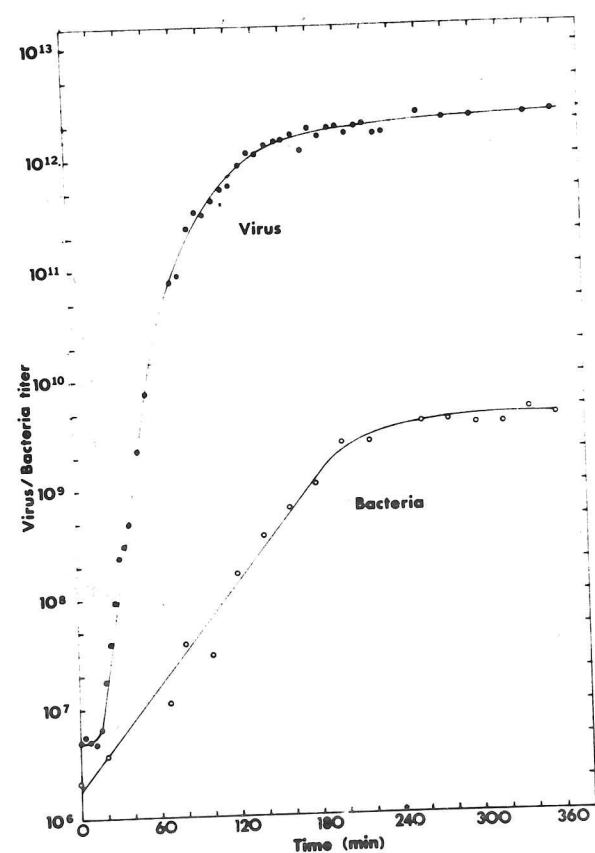
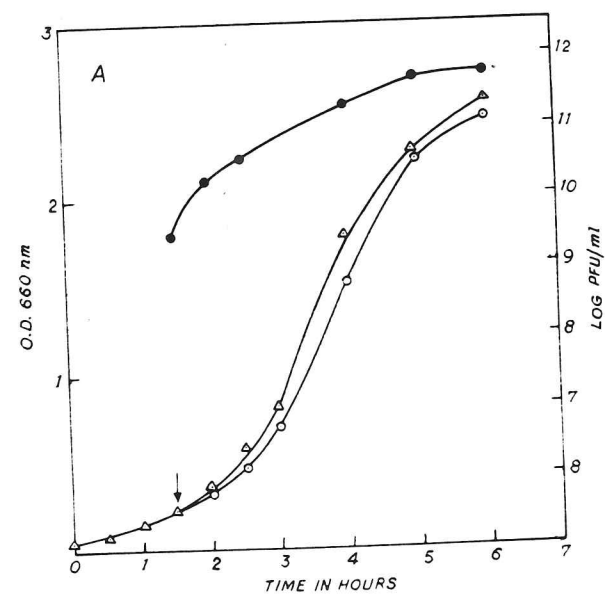


Figure 3.2. The biosynthesis of bacteriophage fd as presented in the literature. The graphs are reproduced directly from: (above) Roy & Mitra (1970a); (below) Marvin & Hohn (1969). In both cases the rate of release of virions (solid circles) appears to decrease after about two hours of culturing, even before the cells (open circles) have reached stationary phase. These reported data appear to contradict the results presented in Figure 3.1.

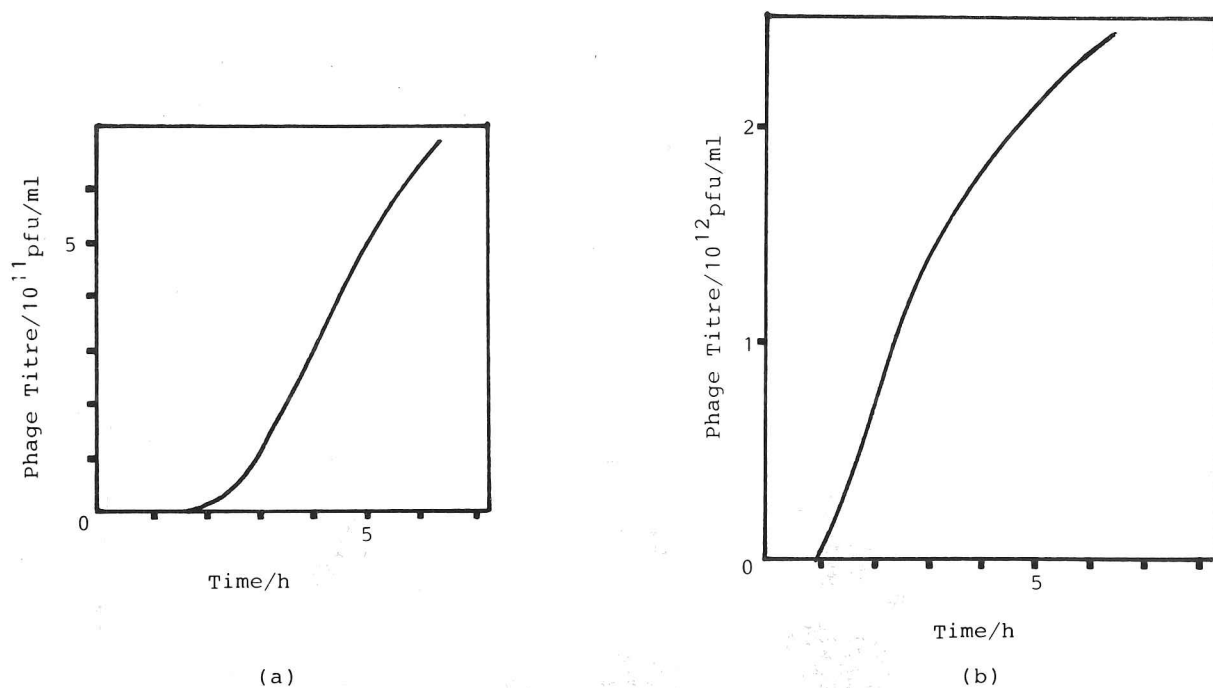


Figure 3.3. The biosynthesis of bacteriophage fd: the resolution of the dilemma. The data on phage release reported in the literature, re-presented in Figure 3.2, were re-plotted on linear axes, (a) Roy & Mitra (1970a), (b) Marvin & Hohn (1969). It is now obvious that phage release does not decelerate sharply after two hours of culturing. The biphasic phage release curves visible in Figure 3.2 are due simply to the semi-log axes on which the results were plotted.

It is clear that there is no sudden decline in phage biosynthesis in either case, phages being produced almost as rapidly at 6 h as near the beginning of infection. In MTPA this rate is about 3 mg phage per litre culture per hour. The apparently biphasic release curve was therefore an artefact of the way in which the data were originally presented.

The primary conclusion which may be drawn from the above is that the culture remains competent to synthesise phages during stationary phase. This implies that the enzyme activities involved in DNA replication and protein synthesis are not limiting phage production. The idea that restricted quantities of polymerases would allow host cell DNA to outcompete viral DNA in replication, thus quashing infection, is rendered unlikely. There is also some circumstantial evidence that this is not so, for phage biosynthesis involves only a small proportion of host cell metabolism and is not likely to compete significantly for these enzymes; indeed, the host cell replicative cycle is hardly affected by infection (Salivar *et al.*, 1964). The practical consequence is that a given culture of cells can continue to generate phage for much longer than the culture period usually quoted in the literature (2 h), and thus can give several-fold higher yields of phage.

Taking the viewpoint of systems control, either the rates of DNA replication, assembly and release are stringently set so that intermediates do not accumulate behind a slow step and kill the cell, or there is some regulation of the system to the same effect. In biological systems the latter is almost invariably the case. Were infection strongly phasic - as was thought previously - such regulation would be relatively unimportant in the short length of the infective cycle.

The control point cannot be at assembly, for then DNA would accumulate and swamp the cell. It has been shown that in longer periods of culture the replicative intermediates are depleted from their log-phase levels (Lerner & Model, 1981), indicating that once DNA is committed to replication and assembly, the process itself is not rate-limiting. The limiting factor must therefore be the initial commitment of (+) strand DNA to replication - by complexing with SSB - or assembly - by complexing with g5p. As phages are released continuously there must be a balance between these competing DNA-binding proteins, g5p taking most of the DNA but SSB taking enough to keep RF replication in line with cell division. This delicate balance is necessary to maintain infection over prolonged periods of growth. It confers an evolutionary benefit on both virus and host, for the virus can propagate to the daughter cells but the cells themselves are not killed (Hoffman-Berling & Mazé, 1964).

### 3.3 Purification Methods

Relatively few viruses have been isolated to a very high degree of purity. Filamentous bacteriophages are more readily purifiable than many others because they have simple, naked protein capsids. They can thus be processed by techniques developed for large protein aggregates, such as precipitations. The nucleic acid increases the density of virions, and the elongated morphology affects the sedimentation characteristics, so that density or velocity sedimentation may also be utilised. Both approaches have been applied to the isolation of viruses from cultures of infected cells. The cells are invariably removed by low speed centrifugation beforehand to leave a phage-bearing supernatant.

Precipitation of phages has been performed with polyethylene glycol 6000 (PEG 6000), a very gentle precipitant compared with ammonium sulphate (Yamamoto *et al.*, 1970). The principle of its action has been studied, and appears to depend on the competitive sequestration of water. The long chain polymer is highly polar and orders a large quantity of water around itself. The filaments of phage similarly order water around their negatively charged surfaces. It is thought that addition of PEG 6000 to a phage bearing supernatant removes so much free water that the phage aggregates into a hydrated gel structure, which is readily sedimentable. There seems to be no direct interaction of PEG with phages, for precipitates recovered after centrifugation contained only as much PEG as would be expected if it were trapped in the interstices of the pellet (Knoll & Herman, 1983; Atha & Ingham, 1981). The alternative approach has been differential centrifugation. This process is usually repeated several times to give a reasonable degree of purification. It has the notional advantage of not introducing further chemicals into the system.

The relative efficiencies of these methods were compared. A culture of *E.coli* K38 cells in 100 ml MTPA medium was prepared, infected with fd phage moi 100 when  $A_{680}$  reached 0.16, and cultured for 2 h (details in section 2.3.1). The cells were removed by centrifugation at 10,000 x g for 10 mins at 4°C, and the supernatant divided into two equal portions.

One portion was brought to 5% PEG 6000, 0.5 M sodium chloride and left overnight at 4°C. A misty suspension was obtained which was



centrifuged at 10,000 x g for 90 min at 4°C. The white pellet so obtained was resuspended in 10 ml NET-buffer and reprecipitated. Samples were removed throughout and analysed by plaque assay. The second portion was centrifuged at 15,000 x g for 20 min at 20°C to pellet debris, then the supernatant at 170,000 x g for 120 min at 20°C. The pellet was resuspended in NET-buffer and the process repeated. Plaque assays showed that the differential centrifugation gave a recovery of approximately 45% of infectious phage particles per cycle. Since several rounds of this might be necessary to get the material pure, the overall yield would be very low. In contrast PEG purification gave recoveries of around 85% per step, allowing much greater efficiency of isolation. This falls off when less than 10 mg/l of phage is present in the starting suspension, but a further precipitation with 50% more PEG recovers most of the missing product. This is in agreement with the observations of Yamamoto et al. (1970), who showed that recovery could be sharply dependent on phage concentration.

Some modern protocols (Lin et al., 1980) call for two such PEG precipitations followed by a direct sedimentation at high g-force, whereas older protocols appeared to achieve purity with fewer steps. To determine whether the extra stages are necessary, phages were purified from a radiolabelled culture of cells. The culture procedure was as given above, except that 0.02 mCi of [<sup>3</sup>H]-arginine were added to the culture with the phages. The proteins were thus radiolabelled, and purification was followed in terms of pfu/1000 dpm of label.

The results showed that both precipitation steps are necessary, each giving a significant purification factor (Table 3.1). The second thus can remove non-phage material left by the first, probably because the precipitation is more selective in the absence of the variable amounts of salts and proteins present in the crude supernatant. The subsequent sedimentation does give some purification but essentially removes traces of PEG from the phages.

The concentrations of phages from such preparations estimated from protein staining in gels (relative to standard samples) were approximately threefold higher than those given by plaque assays. This suggested that phages were being inactivated at some stage. Resuspension of the phage pellet after sedimentation at  $170,000 \times g$  was of necessity vigorous and thus suspect: the shearing involved in the process may have been breaking filamentous capsids. A further culture of 100 ml was prepared and the phages purified. Samples of the suspension were taken at intervals during the resuspension of the pellet and analysed by plaque assay. Vortexing for 10 mins with 1 ml NET-buffer left visible flecks of pellet, and  $3.4 \times 10^{13}$  pfu in suspension. Stirring for 2 h on a magnetic stirrer (PTFE spin-bar, 20 Hz) clarified the material, the detectable pfu rising to  $5.5 \times 10^{13}$ . Extended stirring caused an opalescence to develop in the suspension, with a drop in pfu to  $3.3 \times 10^{13}$  over 24 h. After three days of stirring a floccular white precipitate had formed with very few pfu left in the preparation. The virus was being inactivated by prolonged resuspension, presumably by the shear forces of mixing and turbulence.

Table 3.1. Purification of Phages

Purification step	% Yield	Purification factor
1. PEG precipitation	83	48
2. PEG precipitation	85	3.0
3. Sedimentation	90	1.6
4. CsCl gradient	c.90	variable
5. Sedimentation	92	1

The purification procedure outlined in section 3.3 was performed on 50 ml of cell-free supernatant from an fd-infected culture of E.coli K38. The culture was grown in MTPA medium with 0.02 mCi of [<sup>3</sup>H]-arginine, and samples removed between subsequent stages of purification for plaque assay (section 2.2) and scintillation counting (section 2.9). Purification was followed as the change in pfu/dpm recovered at each stage.

These aspects of the purification process show how important it is not only to compare available methodologies critically, but also to check the results in different ways which can be correlated. After this initial purification, the next step in the case of Ff phages is usually a caesium chloride equilibrium density gradient centrifugation, phages banding at a density intermediate between that of protein and that of DNA. This was performed in the 5 ml tubes of a Beckman SW 50.1 rotor. The resuspended phages were diluted to 4 ml with NET-buffer, then 2 g caesium chloride were dissolved in the suspension to give a final volume of 5 ml. Centrifugation was carried out at 35,000 rpm for 18 h at 20°C, after which the gradient was cut into ten to twelve fractions. Assuming that all the phages loaded onto the gradient banded in at most two fractions - normally the case - samples of suitable volume were taken for visualisation by agarose gel electrophoresis (details in section 2.7). The phage bearing fractions thus identified were pooled and diluted 10-fold with NET-buffer, then centrifuged at 170,000 x g for 5 h at 20°C. The pellet was resuspended in an appropriate volume of buffer, normally 0.05 ml NET.

### 3.4 Protein Analysis

Denaturing polyacrylamide gel electrophoresis in the presence of SDS has been the method of choice for separating and identifying phage coat proteins (Simons et al., 1979). Approximate quantification can be achieved by staining but, if the proteins are radiolabelled, slicing of the gel followed by scintillation counting gives more accurate results.

The problem of denaturation by detergents is fundamental to the analysis of phage capsid proteins, since they have evolved to be extremely resistant to denaturation in the function of protecting the DNA. Failure to disaggregate the capsids fully, and to denature the protein subunits, causes material to migrate in an electrophoretic gel behind the main band as a "streak". This loss of resolution may be avoided by ensuring denaturing conditions in the gel which are adequate to deal with the amount of phage material loaded (Newman et al., 1977). Assuming these conditions are fulfilled the order of protein bands on a polyacrylamide gel for phage fd should be (top, least mobile) g3p, g6p, g8p, g7p.g9p (bottom, most mobile).

In the early stages of this study three electrophoretic systems were used without preference: those described by Laemmli (1970), Wickner et al. (1978) and Simons et al. (1979). An overall picture of the relative performances of these systems, based on several gels, was thus obtained. A membrane-free bacterial lysate prepared from E.coli cells was used to check resolution as it shows a series of finely spaced bands over a wide range of relative molecular masses. Maximum possible loading of fd (more difficult to denature than Pf1) was also estimated (Table 3.2).

Both Laemmli and Wickner gel systems showed good separation of bands in the high relative molecular mass range, though the Wickner system showed better resolution in the lower ranges. This is probably due to the fact that the Laemmli gel is run in 8 M urea, at which concentration the urea displaces some SDS from peptide linkages. The resulting

Table 3.2. Comparison of Polyacrylamide Gel Systems

	Laemmli	Wickner	Simons
Acrylamide	16.8%	15%	15%
Bisacrylamide	0.45%	0.058%	0.8%
Urea	8 M	6 M	8 M
SDS	0.1%	0.1%	0.1%
Tris buffer	0.375 M	0.33 M	0.5 M
pH	8.8	8.7	8.9
Urea in running buffer	8 M	-	6 M
Resolution near origin	Good	Average	Poor
Resolution of g8p band	Poor	Good	Good
Max. phage loading/track	1/2 $\mu$ g	1 $\mu$ g	3 $\mu$ g

heterogeneity of charge is reflected in the varying electrophoretic mobilities. The Wickner gel also allowed higher loadings of phage material, though this was probably due to the more open structure of the gel. The Simons gel system was far superior to either of these, both in the compaction of the g8p band and the amount of phage that could be loaded without streaking. This was ascribed to the use of 6 M urea, promoting denaturation without affecting the binding of SDS, and to the stronger buffering system. The significance of very strong buffering for gels used with phage proteins was only realised when a Simons gel was prepared with 0.25 M Tris.HCl buffer instead of the normal 0.5 M buffer. Very bad streaking was observed even with sub-microgramme quantities of phage.

Having adopted the Simons gel system as the best for analysing capsid proteins, attempts were made to enhance its performance. Linear gradient gels were attempted from 5 to 25% acrylamide. In theory the upper zone allows the wider separation of high molecular mass species, while the lower zone stringently separates smaller species with minor size differences. Unfortunately, the Simons gel mix has a crosslinking ratio well off the ideal for drying down for fluorography. Gels more concentrated than 15% were too inelastic to survive the process and shattered, even when the procedure was attempted without heating. Gradient gels from 5 to 15% acrylamide were found to offer no significant advantages over the isocratic 15% gels, which were therefore used throughout the subsequent course of this work.

The most sensitive staining method currently available is the silver-stain. After comparison of three of the systems available, the method of Morrissey et al. (1981) was adopted on the grounds of sensitivity and low background staining. The g8ps of the fd and Pf1 viruses were thus identified on polyacrylamide gels as the major species present, with the mobility expected for completely denatured proteins of their sizes. Calibration of each gel for relative molecular mass ( $M_r$ ) was performed by running standards (Section 2.4.2). Each g8p has a characteristic appearance: that of Pf1 is a fairly well-defined broad grey band whereas that of fd is a slightly diffuse brown streak. Both proteins have an apolar central section, but that of fd is the more hydrophobic (Torbet, 1979; Bailey et al., 1977; Nakashima et al., 1974c). Since the hydrophobic interaction is the primary agent of intersubunit association, it might be expected that g8p would be more difficult to disaggregate in fd than Pf1 and thus give rise to streaking. No suggestion can be made at present to explain the colour difference.

Artefact bands at apparent  $M_r$  68,000 and 54,000 have been observed regularly in the Simons gel system. It is reported that these are caused by the inclusion of 2-mercaptoethanol in the loading buffer (Tasheva & Desser, 1983). Faint bands, corresponding to those reported, were occasionally seen (Fig. 3.4). They were present even in tracks given loading buffer only, and were observed with different, fresh batches of the buffer. It is difficult to see how gradual changes in the running conditions could give rise to bands of such precise and repeatable mobility. On the other hand polymerisation of the constituents of the





Figure 3.4. Non-protein features of silver-stained polyacrylamide gels. The stainable artefact bands migrating with apparent  $M_r$  68,000 and 54,000 (indicated), observed by Tasheva & Desser (1983), were occasionally evident.

buffer during boiling is rather unlikely to give a product of such well defined size. Of course the bands could well be reduced components of the buffer, running at low mobility due to their small net native charge.

Although silver-stain is a sensitive method for detecting proteins in gels, it would be difficult, by this method, to show up the minor protein components of capsids that have been reported (Simons et al., 1981). The problem is that the major component of the capsids, g8p, is present in several hundred-fold molar excess. Consequently, if g8p is overloading the gel there is still too little of these minor components to stain. The g3p is a larger molecule, however, and contains more binding sites for the silver, making it easier to detect. A heavily loaded Simons gel of phages fd and Pf1 is shown in Figure 3.5. The g3p bands are also visible, at low mobility, in each track, as expected in fd (Henry & Pratt, 1969) but novel in the case of Pf1. A fainter band migrating faster than g8p was observed in both fd and Pf1 tracks. This is the position in which g7p.g9p have been reported for fd (Lin et al., 1980; Simons et al., 1979).

These preliminary results with silver-staining appear to show that phage Pf1 has a g3p, and perhaps counterparts of the smaller minor coat proteins found in phage fd. In the next chapter a more specific search is described, in which selective incorporation of radiolabelled amino-acids was used to label potential minor coat proteins.

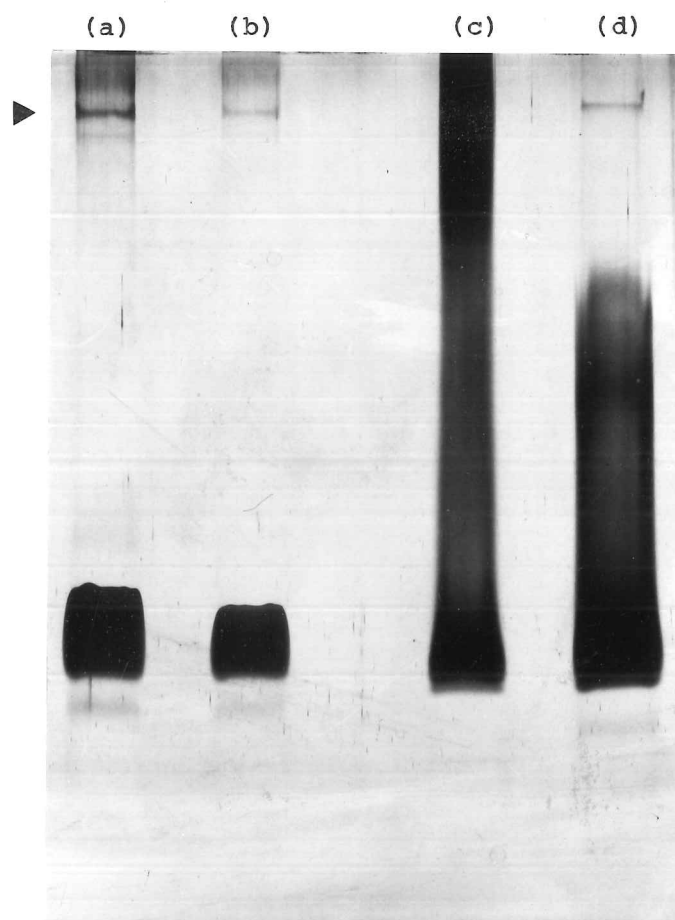


Figure 3.5. A heavily loaded, silver-stained, Simons polyacrylamide gel of bacteriophages Pf1 (a,b) and fd (c,d). The gel was prepared and run as described in section 2.4. The tracks were loaded with (a) 10  $\mu$ g Pf1 phage, (b) 5  $\mu$ g Pf1 phage, (c) 10  $\mu$ g fd phage, (d) 5  $\mu$ g fd phage. The g3p is arrowed, near the origin. Dense bands of g8p are visible in each track. The g3p of fd is visible in (d) but obscured by the streaking of g8p in (c). A corresponding band in the Pf1 tracks indicates the presence of the hitherto unreported Pf1 g3p. This, and the faint band migrating ahead of the very broad g8p suggest that Pf1 capsids may have a complement of minor proteins.

## CHAPTER 4

### THE PROTEINS OF THE PHAGE CAPSIDS

#### 4.1 Selective Radiolabelling of Capsid Proteins

As discussed in the previous chapter, the phage fd capsid proteins are small polypeptides, with the exception of the g3p. On polyacrylamide gels they may be obscured by the vast excess of major coat protein. A method of specifically radiolabelling these proteins has been described, based on the absence of certain amino-acids from the g8p (Lin et al., 1980). Thus by radiolabelling with an amino-acid that is not represented in the major coat protein, only the minor proteins will be labelled and they can be analysed by electrophoresis and fluorography. A random search need not be carried out for these purposes as the amino-acid sequence of the g8p is known (Bailey et al., 1977; Nakashima & Konigsberg, 1974; Asbeck et al., 1969). One amino-acid which should label the minor proteins without labelling the g8p, is arginine (Table 4.1). This can be incorporated into phages by the addition of radiolabelled arginine to cultures of infected cells. Arginine is readily available in tritiated form, which has the benefits of a long half-life and a high specific radioactivity

The experimental system must meet two requirements. First, the specific radioactivity of the virions must be sufficient to give clear fluorographs at reasonable gel loadings. Pilot trials showed that about  $6 \times 10^6$  dpm/mg would be adequate. Second, it is very important that tritium should not enter other amino-acids (and hence g8p) via interconversions under host cell metabolism. The latter condition may be achieved by adding small amounts of arginine at high specific

Table 4.1. Selected Open Reading Frames of the fd DNA:

Amino-acid Compositions (data of Beck et al., 1978)

Gene:	3	6	7	9	8 <sup>*</sup>	5
Ala (A)	28	7	4	1	10	4
Arg (R)	9	1	1	2	0	4
Asn (N)	29	2	0	0	0	2
Asp (D)	24	5	2	0	3	3
Cys (C)	8	1	1	1	0	1
Gln (Q)	15	4	4	0	1	6
Glu (E)	26	0	1	1	2	4
Gly (G)	68	8	3	2	4	7
His (H)	2	0	0	0	0	1
Ile (I)	9	13	5	1	4	4
Leu (L)	19	24	2	4	2	10
Lys (K)	14	7	0	0	5	6
Met (M)	7	1	2	2	1	2
Phe (F)	20	10	2	3	3	3
Pro (P)	25	5	0	0	1	6
Ser (S)	32	8	1	6	4	7
Thr (T)	26	2	1	3	3	4
Trp (W)	4	1	0	1	1	0
Tyr (Y)	21	4	1	2	2	5
Val (V)	20	9	3	3	4	8
Total	406	112	33	32	50	87

\* Based on mature g8p, after signal peptide has been removed.

radioactivity to arginine-limited cell cultures. The labelled amino-acid should be taken up rapidly and not wasted by conversion to other metabolites.

An obvious approach was the use of auxotrophs in arginine since the added radiolabelled amino-acid could not be diluted by unlabelled arginine synthesised within the cells. Three strains of E.coli which were auxotrophic in arginine and bore sex-pili ( $F^+$  or HFr types) were kindly supplied by the E.coli Genetic Stock Center, Yale: P4X SB147, MA 1013 and AT 12-55. These were prepared as streaks on Luria agar, and their auxotrophic stringency was tested by culturing in MTPA medium with and without arginine. Being a defined medium, MTPA allows the exclusion of single amino-acids required in these experiments. All three strains were found to have a high stringency, growth being inhibited to at least 95% in the absence of arginine (Table 4.2). Most stringent was AT 12-55, MA 1013 being intermediate and P4X SB167 least stringent.

Since no other source of arginine is available to these cells, the specific radioactivity of the arginine added exogenously may be related directly to the specific radioactivity of the phages produced (through the reported stoichiometries of the minor proteins, and their component amino-acids). Given the cost and specific radioactivity of the [ $^3H$ ]-arginine available it was decided that no more than  $10^{-5}$  M arginine could be allowed in the growth medium (after dilution with unlabelled arginine), or the specific radioactivity of the product would be too low. Cultures of each strain were prepared in MTPA with this

Table 4.2. Growth of Arginine Auxotrophic Strains of E.coli

Strain	Arginine			Phages
	concentration: 3 $\mu$ M      12 $\mu$ M      1 mM			released
				pfu/ml/h
P4X SB167	0.030	0.108	0.651	$1.8 \times 10^{10}$
MA 1013	0.041	0.163	1.182	$1.2 \times 10^{10}$
AT 12-55	0.010	0.111	1.083	$0.4 \times 10^{10}$
K38(isotroph)	1.272	1.129	1.026	$1.0 \times 10^{11}$

The table shows the  $A_{680}$  of cell cultures grown up in MTPA medium for 12 h. Each culture was started at  $A_{680}$  0.010 and infected with fd phage at moi 100. At the end of the culture period the absorbance was measured; in the case of the 12  $\mu$ M cultures the titre of pfu was determined by plaque assay.



concentration of arginine. They were infected at  $A_{680}$  0.04 with phage fd at moi 100 and samples withdrawn at intervals for plaque assay. A reference culture of E.coli K38 was also prepared in the same medium. The rate of phage production by the mutants was very low compared with this non-auxotrophic strain. The data in Table 4.2 relate to the total phage synthesis under these mock radiolabelling conditions over 12 h of culture. With such a low phage production by mutants the total recovery of radiolabel in phages would be very poor. The K38 strain on the other hand produced phages at a reasonable rate. Thus, if host cell metabolism does not contribute too much to the intracellular arginine pool, it appeared that the wild-type would be a better host for radiolabelling experiments.

As was shown in the previous chapter, phage biosynthesis continues for many hours after infection. To examine whether the phages are produced at a constant specific radioactivity over a long period, a culture of 100 ml E.coli K38 was prepared in MTPA(-Arg) and infected at moi 100 with phage fd in mid log-phase growth (section 2.3.1). Radiolabel was added as 0.05 mCi of [ $^3\text{H}$ ]-arginine at 28.8 Ci/mmol. Samples of 0.5 ml were withdrawn every hour, from which 0.1 ml was used to assay  $A_{680}$ . The remainder was centrifuged at 10,000 x g for 10 min to remove the cells. The supernatant was decanted, and 0.1 ml reserved for plaque assay, the rest being brought to 5% PEG 6000 and 0.5 M NaCl by the addition of an appropriate quantity of 25% PEG 6000, 2.5 M NaCl. The phages were allowed to precipitate at 4°C overnight, then pelleted by centrifugation at 10,000 x g for 30 min at 4°C in an Eppendorf

centrifuge. The first precipitate was resuspended and reprecipitated to purify the phages, the product being subjected to scintillation counting and plaque assay (to allow correction for losses in the purification). In each case, the supernatant of the first precipitation was further cleared by the addition of 25% PEG, 2.5 M NaCl to 10% PEG, in order to remove residual phages. Scintillation counting of these samples showed the residual level of radiolabel in the medium. This stabilised within 30 min of the addition of radiolabel at 60 nCi/ml - corresponding to an arginine concentration of 2 nM. Arginine transport is therefore quite rapid, and the residual 2 nM arginine may well correspond with the threshold of the arginine transport system. Such a threshold is comparable with those of other transporters in the bacterial membrane.

Phage biosynthesis was, as is now to be expected, linear with respect to time (Fig 4.1). Radiolabel in phage was not released in a manner detectably different from linear, indicating that the pool of available arginine within the cell is much larger than that needed for phage biosynthesis, and is therefore not depleted of radiolabel during the course of the culture. Phages were produced throughout at a specific radioactivity of 120,000 dpm/mg, as determined by scintillation counting (section 2.9) and plaque assay (section 2.2). Thus it is possible to run radiolabelled cultures for much longer times than had been used until now (Lin et al., 1980; Simons et al., 1979). Neither phage production nor specific radioactivity tail off in stationary phase cultures. Up to 9 h culturing may be used, generating 4-5 times as much product as the protocols cited in the literature. This single modification of the

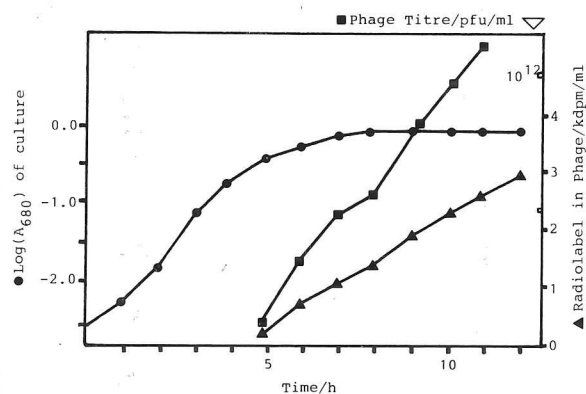


Figure 4.1. The biosynthesis of selectively radiolabelled bacteriophage fd. A culture of 100 ml *E.coli* K38 in MTPA(-Arg) medium was infected with phage fd at moi 100 and 0.05 mCi of [<sup>3</sup>H]-arginine added. Samples were taken every hour: the A<sub>680</sub> was determined to allow the progress of the culture to be followed (circles); the phage titre was determined by plaque assay (squares); and the radioactivity recovered in phage was determined by scintillation counting (triangles). Both phage biosynthesis and the release of radiolabel in viral particles was approximately linear with respect to time for several hours after the culture had reached stationary phase.

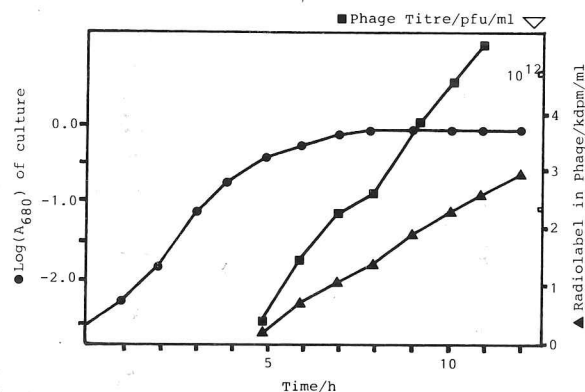


Figure 4.1. The biosynthesis of selectively radiolabelled bacteriophage fd. A culture of 100 ml *E.coli* K38 in MTPA(-Arg) medium was infected with phage fd at moi 100 and 0.05 mCi of [ $^3$ H]-arginine added. Samples were taken every hour: the  $A_{680}$  was determined to allow the progress of the culture to be followed (circles); the phage titre was determined by plaque assay (squares); and the radioactivity recovered in phage was determined by scintillation counting (triangles). Both phage biosynthesis and the release of radiolabel in viral particles was approximately linear with respect to time for several hours after the culture had reached stationary phase.

protocol gives a 500% increase in phage yield from a given amount of label, or conversely, allows a given amount of radiolabelled phage to be prepared for 80% less cost.

Use of the K38 strain is subject to the conditions outlined previously - the highest concentration of arginine in the growth medium that can be provided is  $10^{-5}$  M and there must be minimal host cell synthesis of this amino-acid. A brief trial was performed to determine whether the additional unlabelled arginine was necessary at all with this strain. Two parallel cultures of E.coli K38 were grown in 50 ml MTPA each, one with neat radiolabelled arginine (0.02 mCi at 18.6 Ci/mmol giving 80 nM arginine), the other with additional arginine to 12  $\mu$ M. The extra arginine might enhance phage production without significantly diluting the radiolabel. The total recovery of radiolabel from the two cultures was similar (Table 4.3). The additional arginine did increase the rate of phage biosynthesis, but diluted the radiolabel to about the same extent. The inclusion of this level of arginine in the medium was therefore deemed unnecessary, indeed detrimental since it reduced the specific radioactivity of the product.

The culture without extra arginine produced virus at a specific radioactivity of  $7 \times 10^5$  dpm/mg as determined by scintillation counting (section 2.9) and plaque assay (section 2.2). This is well below the desired value of about  $6 \times 10^6$  dpm/mg. Further cultures were attempted with successively higher concentrations of radiolabelled arginine. The phages were isolated from each and the specific radioactivity assessed

Table 4.3. Phage Production Under Various Radiolabelling Conditions

[ <sup>3</sup> H]-arginine supplied at	18.6	18.6	28.8	28.8	Ci/mmol
Level of radiolabelling	1.5	1.5	10	100	μCi/ml
Arginine concentration	12	0.08	0.35	3.5	μM
Phage production	3.4	0.4	3.2	3.3	μg/ml/h
Radioactivity recovered	2.2	1.8	60	120	kdpm/ml
Specific radioactivity	0.1	0.7	3.2	6.0	Mdpm/mg

fd phages were purified from 6 h cultures of E.coli K38 in MTPA medium with various amounts of [<sup>3</sup>H]-arginine added. The standard phage culturing and purification methods were used (section 2.3) and samples of the recovered phages analysed by plaque assay and scintillation counting.

Table 4.4. Selected Open Reading Frames of the Pf1 DNA:

Amino-acid Compositions (data of N.J. Short, pers. comm.)

Gene:	3	6	7	9	8 <sup>*</sup>	5
Ala (A)	38	20	5	11	7	18
Arg (R)	17	13	1	2	1	6
Asn (N)	21	0	0	1	0	4
Asp (D)	32	7	1	1	3	7
Cys (C)	17	4	0	2	0	1
Gln (Q)	8	6	0	3	2	14
Glu (E)	19	3	0	4	0	7
Gly (G)	63	11	4	11	7	11
His (H)	0	3	0	0	0	0
Ile (I)	13	11	1	4	6	7
Leu (L)	30	12	6	8	4	7
Lys (K)	22	2	0	2	2	8
Met (M)	6	5	1	2	2	3
Phe (F)	8	8	1	4	0	5
Pro (P)	32	11	0	5	0	13
Ser (S)	36	12	2	5	3	6
Thr (T)	33	6	2	4	2	12
Trp (W)	6	3	2	3	0	1
Tyr (Y)	15	4	2	2	2	3
Val (V)	22	6	2	9	5	11
Total	438	147	30	83	46	144

\* Based on mature g8p, after signal peptide has been removed.

by scintillation counting and plaque assay. Cultures were grown for 6 h at levels of 1.5, 10 and 100  $\mu\text{Ci/ml}$  [ $^3\text{H}$ ]-arginine in MTPA(-Arg). These gave progressively higher specific radioactivities of virus, as shown in Table 4.3. Labelling at the level of 1 mCi per 10 ml medium (3.5  $\mu\text{M}$  in arginine) gave a product of sufficient specific radioactivity for the purposes of electrophoresis and fluorography, and a good yield of phage at reasonable cost.

#### 4.2 The Minor Proteins of fd Capsids

The culturing conditions refined in the previous section were then applied to the preparation of a larger quantity of specifically radiolabelled phage, for electrophoretic analysis. A culture of E.coli K38 in 10 ml MTPA(-Arg) was infected in mid-log-phase growth with fd phages and radiolabelled with 1 mCi [ $^3\text{H}$ ]-arginine at 28.8 Ci/mmol. The phages were isolated after 9 h of culture and purified by the standard protocol. Since complete radiochemical purity was required, further purification was undertaken. Ff phages are stable in high concentrations of salt, so equilibrium density centrifugation in caesium chloride was performed. A self-forming gradient of 5 ml was prepared as 0.4 mg/ml caesium chloride in TE-buffer and centrifuged at 110,000  $\times g$  for 18 h at 20°C. The gradient was fractionated, and the phage localised by agarose gel electrophoresis (section 2.7). The exquisite separation of this step is shown in Figure 4.2. The relevant fraction was diluted 10-fold in NET-buffer and sedimented at 170,000  $\times g$  for 5 h at 20°C. Phages were resuspended in 0.1 ml NET-buffer.



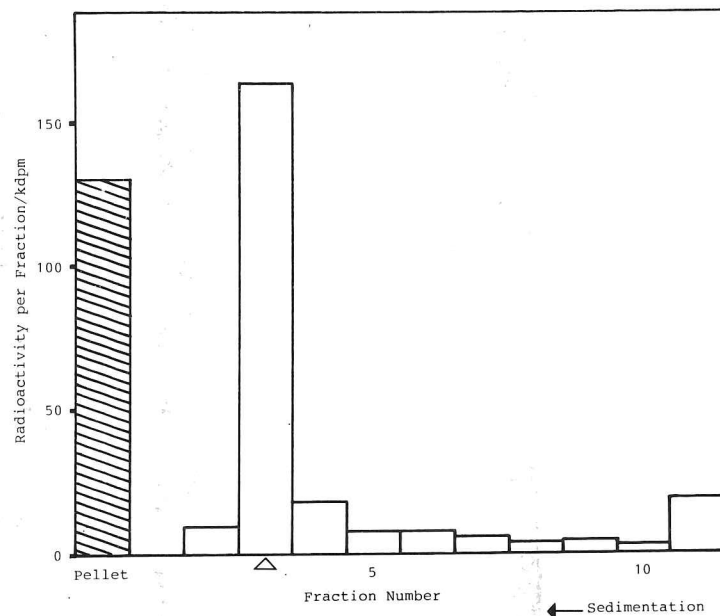


Figure 4.2. The radioactivity profile of a caesium chloride gradient of phage fd. Sedimentation was from right to left. The arrowhead indicates the fraction which contained the viral DNA. It is well separated from the radioactivity associated with the DNA, which reached the bottom of the gradient (shaded), and from contaminating proteins which float at their buoyant density near the top of the gradient.

A total of  $2 \times 10^6$  dpm were recovered. Samples containing 25,000 dpm were analysed by polyacrylamide gel electrophoresis according to the Simons protocol (section 2.4). Reference tracks of unlabelled phages and size markers were run in one half of the gel, which was then silver-stained to check that the electrophoresis had been successful and to calibrate the gel. The other half of the gel, with the radiolabelled track, was impregnated with PPO (section 2.4.4) which was then precipitated in situ. The gel was dried down and exposed to X-ray film in a light-proof cassette at  $-20^{\circ}\text{C}$ . A parallel experiment was done with [ $^3\text{H}$ ]-tryptophan, to allow comparison with a phage sample radiolabelled in an amino-acid not absent from the major coat protein. A culture of fd-infected E.coli K38 was grown in MTPA(-Trp) with 1 mCi of [ $^3\text{H}$ ]-tryptophan at 3.4 Ci/mmol, and the phages isolated as above.

The Trp-labelled phage showed a fluorographic pattern quite like that in silver-stain (Fig. 4.3). A dense g8p band was visible, and a faint g3p band. The g8p did not streak, because very little protein is present in the track and it is therefore adequately denatured and buffered. As with silver-stain, any other minor proteins are not visible, because of their low concentration with respect to g8p. In contrast, the arginine labelled phages show no g8p band at all - confirming that no metabolic leakage of radiolabel has occurred (Fig. 4.3). The g3p was heavily radiolabelled, the band comigrating with the g3p seen in silver-stain. The band migrating just behind g8p, and thus corresponding to a larger protein, is probably g6p. One migrating just ahead of g8p is probably g7p.g9p, for both constituent polypeptides are approximately the same size. This lower band is in the same position ahead of g8p as the faint silver-staining band seen on a gel presented previously (section 3.4).

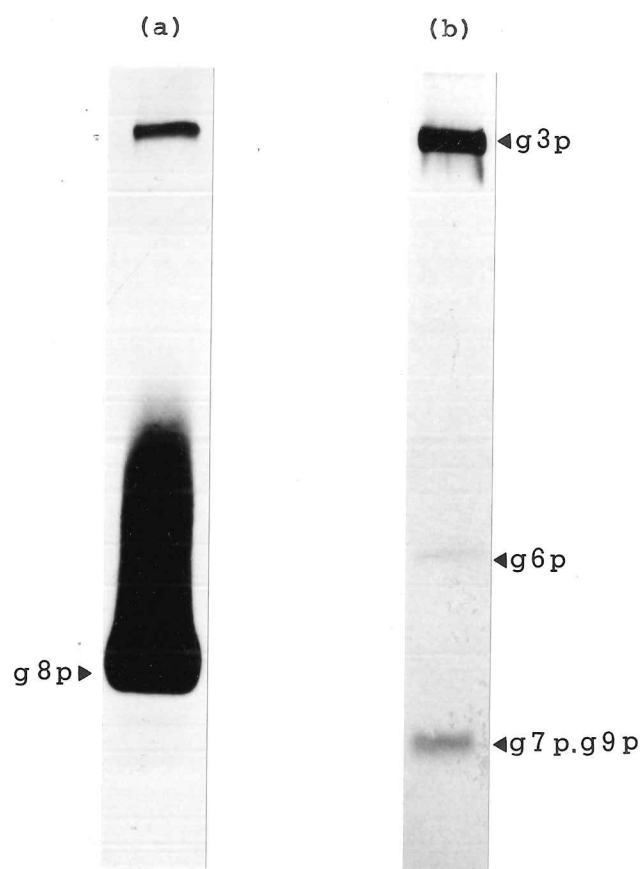


Figure 4.3. Fluorographs of selectively radiolabelled phage fd, analysed by polyacrylamide gel electrophoresis. Track (a) contains fd phages isolated from a 9 h culture of *E.coli* K38 in 10 ml MTPA(-Trp) medium to which had been added 1 mCi of [ $^3\text{H}$ ]-tryptophan. The g8p is clearly visible as a heavy, streaked band; a band of g3p is also visible, but other minor protein components of the capsid cannot be seen. Track (b) contains phages radiolabelled with [ $^3\text{H}$ ]-arginine, and amino-acid not present in g8p, which is therefore invisible. The bands in this track have mobilities and relative intensities consistent with their assignments as (from origin) g3p, g6p and g7p.g9p (the last two comigrating).

The relative molecular masses of the proteins observed were estimated by comparison with the migration of proteins of known  $M_r$  in the same gel. A plot of  $\log (M_r)$  versus migration distance is shown in Figure 4.4. The relative molecular masses of these putative proteins have been calculated (Appendix II) from the DNA sequence (Beck *et al.*, 1978) using a computer program developed by Dr. P.R. Alefounder (pers. comm.). The bands assigned to g3p, g6p, g8p, g7p and g9p show mobilities consistent with this interpretation if compared with the calibration curve. The corresponding genes show that g3p contains 9 arginine residues, g6p 1, and g7p.g9p 3 (Table 4.1). These figures correlate nicely with the observed relative intensities of the bands, g3p being more dense than g7p.g9p which is in turn denser than g6p.

The technique of *in vivo* radiolabelling is a powerful one, capable of exquisite selectivity and high yields if the correct conditions are found. In conjunction with electrophoretic techniques, it is a useful method for analysing the minor protein components of viral capsids, for these often contain the short length proteins prerequisite for amino-acid exclusions to occur. Previously published methods applying the methods to filamentous bacteriophages have been adequately selective (Lin *et al.*, 1980), but extremely inefficient in comparison with the system developed in the course of this work. Consequently, the system described herein is reasonably inexpensive - in contrast with published forms of this methodology. The application of this technique allowed the confirmation of the presence of gene 3, 6, 7 and 9 proteins in the capsids of Ff phages. This is a useful but not unexpected result. The minor capsid proteins migrate with their expected mobilities, and show arginine contents consistent with their genetic identities.

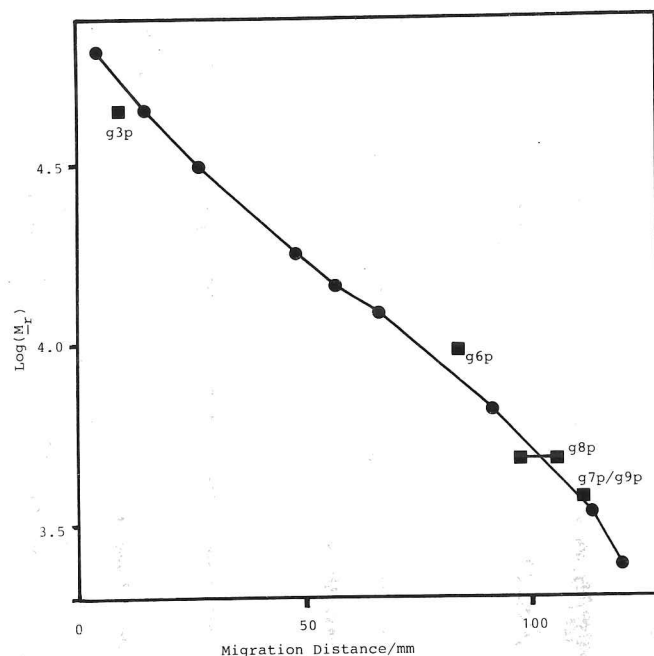


Figure 4.4. The electrophoretic migration distances of minor coat proteins of fd capsids, plotted against the logarithms of their predicted  $M_r$ s. A set of proteins of known  $M_r$ s was used to calibrate the gel (section 2.4), giving the diagonal line. The expected  $M_r$ s of the g3p, g6p, g7p and g9p were derived from the DNA sequence (Beck *et al.*, 1978), and correlate reasonably well with the mobilities of the corresponding bands (marked as squares). The g8p is marked as a bar, reflecting the width of the band normally observed.

#### 4.3 The Class II Phage Pf1

The Ff phages have been demonstrated to contain minor protein components. The question arises whether the capsids of Class II phages, such as Pf1, also contain minor protein components. If they do, detailed consideration of these two sets of proteins could throw more light on their functions than the study of one set on its own.

The study must begin with the consideration of the DNA sequence of Pf1 virus which has been determined, in part, in this laboratory (N.J. Short, pers. comm.). The sequence showed some similarity to Ff phages in the distribution of long and short potential coding sequences, and the relative positions of genes 5 and 8 (N.J. Short, pers. comm.). On the basis of this pattern, putative coding sequences for genes 3, 6, 7 and 9 could be assigned. The amino-acid compositions were obtained by computer aided translation of the relevant parts of the genome (Table 4.5). Six amino-acids were found to be absent from the mature form of the major coat protein: Asn, Cys, His, Phe, Pro, Trp. Asn is also absent from g6p and g9p, Cys is absent from g7p, His from g3p, g7p, g9p, and Pro from g7p. Phe and Trp are present in all four putative proteins. The radiolabelling methods developed for phage fd could therefore be applied to phage Pf1. Little separates the choice of Phe or Trp for this purpose either in cost of radiolabel or potential metabolic leakage of label (unlikely in either case). Trp was chosen because it is present at least twice in every protein molecule considered, whereas Phe is present only once in g7p.

#### 4.4 The Minor Proteins of Pf1 Capsids

The host cell strain of Pf1, Pseudomonas aeruginosum K, was kindly supplied by Dr. D.A. Marvin as a glycerolised culture. The strain was grown in 2xTY medium and streaked out on sterile Luria agar plates (section 2.1). Plaque assays performed with cells from this plate showed a reasonable plaquing efficiency when tested against the Pf1 stock (section 3.1).

Pilot trials of culturing and purification were performed to ensure that the methods developed for fd phage culture and purification were as effective with Pf1. In general this was so, except that Pf1 is more salt-labile and so cannot survive concentrated solutions of caesium chloride intact (Nakashima et al., 1975). For this reason the density gradient step had to be avoided, and three rounds of differential centrifugation substituted. A low speed spin at 15,000 x g for 20 min removed debris, then the supernatant was centrifuged at 170,000 x g to pellet the phages. Recovery was quite acceptable (about 50% overall) compared with the corresponding purification of fd, since the Pf1 pellet was much more readily resuspended. This gave pure phages as judged by polyacrylamide gel electrophoresis (on the Simons system) and silver-staining.

A culture of P.aeruginosum K in 10 ml MTPA(-Trp) was prepared and infected with phages at moi 100. [<sup>3</sup>H]-tryptophan at 3.4 Ci/mmol was supplied as a 1:1 ethanol/water solution, and was transferred to pure

water by drying down under high vacuum, then vortexing the invisible residue with 1 ml sterile water at 37°C. This was added to the culture medium at the time of infection. Owing to the lower growth rate of P.aeruginosum K, cultures were incubated for 12-14 h at 37°C with orbital shaking. The characteristic yellow-green colour of this strain was observed in stationary phase. The standard purification protocol was performed (section 2.3.2), followed by the differential centrifugations described above. The product showed only g8p when analysed by Simons electrophoresis and silver-staining. A high radioactive yield was obtained (about  $10^6$  dpm) from which a phage sample containing 50,000 dpm was loaded onto a Simons gel and analysed by electrophoresis, PPO impregnation and fluorography. This revealed four radioactively labelled protein bands on the gel (Fig. 4.5). From origin to dye front these should correspond with g3p, g6p, g9p and g7p. The intensities are as would be expected, in descending order, g3p (6 Trp), g6p and g9p (3 Trp) and g7p (2 Trp). The uppermost of these comigrates with g3p seen in silver-stain, and may thus be confidently assigned. Note that in contrast with the case in fd, g7p and g9p are of different mobilities.

Protein size markers had been run on the gel and were used to calibrate the plot of  $\log(M_r)$  against migration distance (Fig. 4.6). The electrophoretic mobilities of the observed bands correspond closely with the expected mobilities of g3p, g6p, g9p and g7p, as predicted from the open reading frames of the partial DNA sequence. The putative products of genes 3, 6, 7 and 9 have therefore been found to be present in the Pf1 capsid, migrating under electrophoresis at mobilities consistent with their assignments. This confirms the putative genes as real coding sequences which are expressed in vivo.



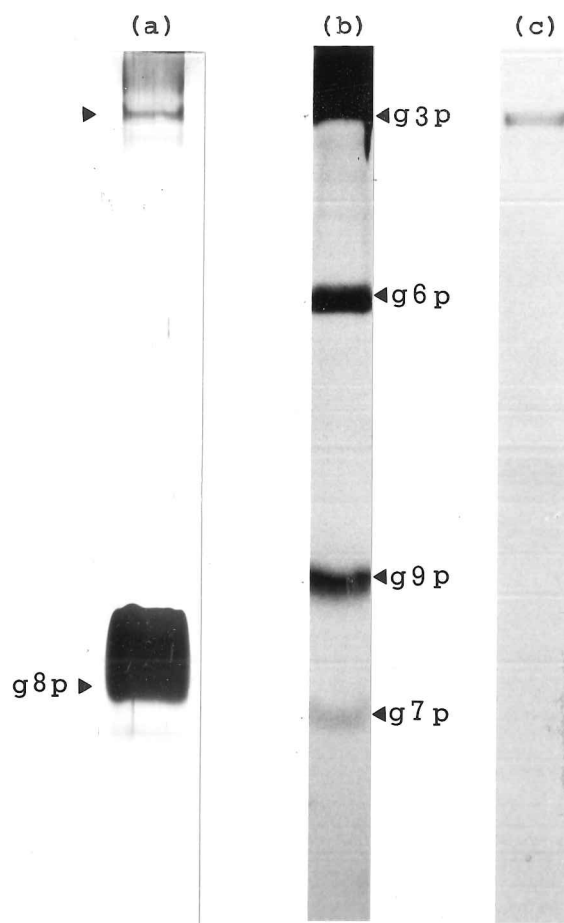


Figure 4.5. Analysis of the proteins of phage Pf1 capsids by electrophoresis. Track (a) was silver-stained, showing a dense band of g8p and a fainter one of g3p (arrowed). Track (b) is the corresponding fluorograph of Pf1 phages isolated from a 12 h culture of *P.aeruginosum* K in 10 ml MTPA(-Trp) medium, to which had been added 1 mCi of [ $^3$ H]-tryptophan. The g8p is not radiolabelled, but four other species are indicated, corresponding with g3p, g6p, g9p and g7p. The g3p is overexposed in (b), but it is visible as a band if the fluorograph is exposed for a shorter time, as in (c), although the other bands are not then visible. Capsids of phage Pf1, therefore, contain a complement of minor proteins genetically comparable with those of fd.

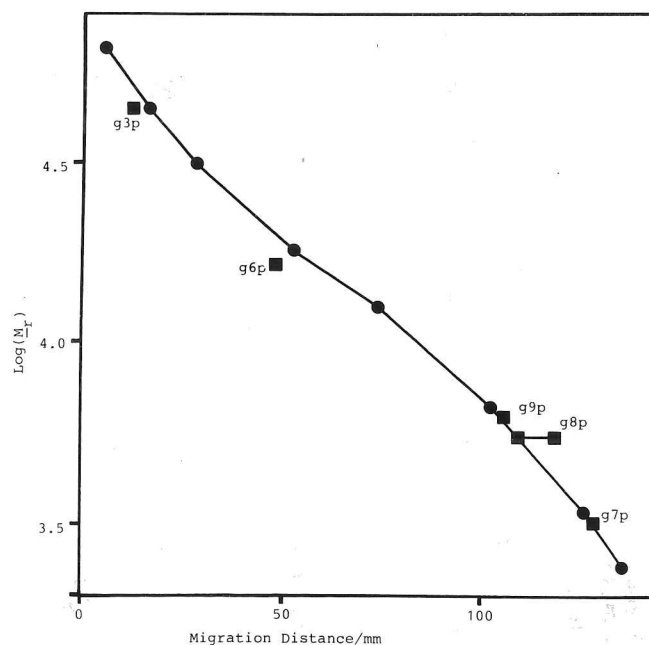


Figure 4.6. The electrophoretic migration distances of the minor coat proteins of Pf1 capsids, plotted against the logarithms of their predicted  $M_r$ s. Protein standards were used as before (Fig. 4.4) to calibrate the system, and are plotted as a line on the graph. The expected  $M_r$ s of the g3p, g6p, g7p and g9p were derived from the DNA sequence (N.J. Short, pers. comm.), and correlate well with the mobilities of the corresponding bands (marked as squares). The g8p is marked as a bar, reflecting the width of the band normally observed.

#### 4.5 The Stoichiometry of the Minor Proteins of Pf1

The stoichiometry of these minor proteins is a point of some interest, related to the symmetry of the virion. In the case of fd, each minor protein appears to be present in about 5 copies per virion (Lin et al., 1980). This may reflect the 5-fold symmetry of the capsid. In the case of Pf1, however, the symmetry is that of a single start helix, so the minor protein stoichiometry is a matter for conjecture (Wiseman & Day, 1977).

Electrophoresis offers a convenient way to separate these minor coat proteins, and radiolabelling allows a route to quantitation via scintillation counting. Transferring the protein from a gel to the scintillant requires three things: the gel must be partitioned into small, equal segments; the gel structure must be disrupted to allow access of scintillant to the radiolabel; and enough radioactivity must be present to allow accurate counting

Approximately 200,000 dpm of [ $^3\text{H}$ ]-Trp labelled phages were loaded onto a Simons gel. Electrophoresis was carried out as normal (section 2.4) and the gel carefully recovered without fixation. The single track of sample was cut out of the gel with a fresh, clean razor blade. This was done with a scissor-like, crimping action as the gel was found (in preliminary trials) to tear if a cutting sweep was employed. Care was taken to ensure two parallel sides were obtained, 1 cm. apart, perpendicular to the dye-front and the stacking gel. The strip was

#### 4.5 The Stoichiometry of the Minor Proteins of Pf1

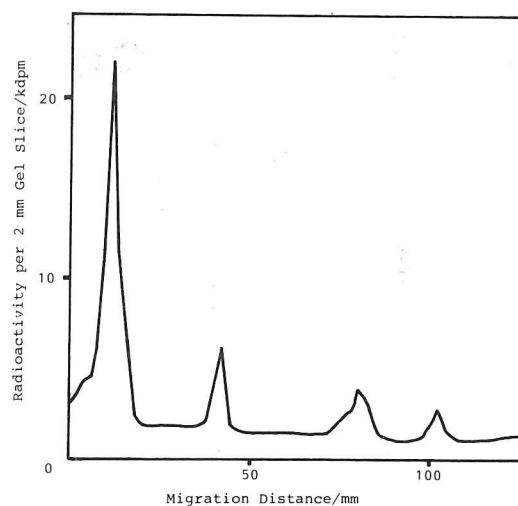
The stoichiometry of these minor proteins is a point of some interest, related to the symmetry of the virion. In the case of fd, each minor protein appears to be present in about 5 copies per virion (Lin et al., 1980). This may reflect the 5-fold symmetry of the capsid. In the case of Pf1, however, the symmetry is that of a single start helix, so the minor protein stoichiometry is a matter for conjecture (Wiseman & Day, 1977).

Electrophoresis offers a convenient way to separate these minor coat proteins, and radiolabelling allows a route to quantitation via scintillation counting. Transferring the protein from a gel to the scintillant requires three things: the gel must be partitioned into small, equal segments; the gel structure must be disrupted to allow access of scintillant to the radiolabel; and enough radioactivity must be present to allow accurate counting

Approximately 200,000 dpm of [<sup>3</sup>H]-Trp labelled phages were loaded onto a Simons gel. Electrophoresis was carried out as normal (section 2.4) and the gel carefully recovered without fixation. The single track of sample was cut out of the gel with a fresh, clean razor blade. This was done with a scissor-like, crimping action as the gel was found (in preliminary trials) to tear if a cutting sweep was employed. Care was taken to ensure two parallel sides were obtained, 1 cm apart, perpendicular to the dye-front and the stacking gel. The strip was

placed on a clean sheet of plastic (Clingfilm) over a sheet of 1 mm graph paper. A razor blade was used to cut the gel into 2 mm slices, each of which was placed in a separate, numbered scintillation vial with a microspatula. Both the razor blade and the spatula were wiped clean with ethanol after each slice. Sixty-three slices were thus obtained; the procedure was completed rapidly as the gel was desiccating, though the lower portions were covered in plastic film to reduce this problem.

The vials were then placed in a large desiccator and dried under vacuum (on a freeze-drier) for 24 h. The gel slices became hard and opaque. To each vial was added 1 ml of 20 vols hydrogen peroxide, ensuring that the slice was at the bottom of each tube and covered with oxidant. After 24 h incubation at 37°C, to disrupt the gels, a PPO based scintillant (section 2.9) was added to each vial (3 ml) and the vials were shaken vigorously to redissolve the precipitated PPO. The scintillation counting was repeated three times, and the mean of each set of values plotted against slice number (Fig. 4.7). This clearly shows the four peaks of radioactivity corresponding with the minor proteins. The ratio of counts in the peaks was determined in two ways: by adding up the total dpm of each peak and correcting for background; and by cutting the peaks out of a graph and weighing them. The results of each method are shown in figure 4.8, together with the mean of the two sets of ratios which is 1:1.25:1.5:3.33 of g7p/g9p/g6p/g3p. Since there are two moles Trp residues per mole g7p, the ratio is better expressed as 2:2.5:3:6.66. The molar ratios of Trp in these proteins are 2:3:3:6 (Table 4.4). The stoichiometry of the minor proteins of Pf1 capsids is therefore approximately 1:1:1:1.



	g7p	g9p	g6p	g3p
Dpm ratio (a)	1.0	1.22	1.47	3.35
Peak ratio (b)	1.0	1.27	1.53	3.31
Mean ratio (c)	1.0	1.25	1.50	3.33

Figure 4.7. The radioactivity profile of a polyacrylamide gel track of [ $^3\text{H}$ ]-tryptophan labelled phage Pf1. The peaks corresponding to g3p, g6p, g9p and g7p (left to right) are visible. The ratio of the dpm in the peaks of g7p/g9p/g6p/g3p was determined directly from the scintillation counting (a), and by plotting the graph, cutting out and weighing the peaks (b). The average ratio (c) indicated a minor protein stoichiometry of about 1:1:1:1 in g3p/g6p/g7p/g9p (in conjunction with the data of Table 4.5).

#### 4.6 Discussion

Extensive homology between the Class I and the Class II phages was first suspected when the partial DNA sequence of Pf1 virus showed a pattern of long and short coding sequences similar to the fd sequence, with genes 5 and 8 in corresponding positions (N.J. Short, pers.comm.). Genes 3, 6, 7 and 9 of fd are thought to code for minor protein components of the virion and the same might be true of genes in corresponding positions on the Pf1 sequence. Radiolabelling in vivo in a selected amino-acid has not only confirmed the presence of these proteins in the capsids of the Class I phage fd, but brought to light the hitherto unknown minor protein components of the Class II phage Pf1. This is indicative of homology in the structures of Class I and Class II capsids. Furthermore, these minor protein components of Pf1 have a 1:1:1:1 stoichiometry, as is the case for fd (Lin et al., 1980).

What then, are the functions of these minor proteins? The function of g3p is well described in fd (Armstrong et al., 1981), but it has never been shown to be the host specific adsorptive principle of Pf1. The g7p.g9p has been implicated in anchoring DNA in the virion, and in initiating assembly (Lopez & Webster, 1984; Lopez & Webster, 1982; Griffith et al., 1981). Yet there is no mechanistic detail beyond these vague generalisations. The obscurity surrounding these proteins is highlighted by their ubiquity, being present in the capsids of both Class I and Class II phages. Since it is canonical in the biological sciences that structure and function are related, more detailed study of the structures of these capsid proteins may shed some light on their incorporation into capsids and their potential functions.

## CHAPTER 5

### SECONDARY STUCTURE PREDICTIONS OF VIRAL PROTEINS



### 5.1 Protein Secondary Structure

The functions of proteins are intimately related to the folding of their polypeptide chains and it is common to find that particular types of folding - supersecondary structures - relate to particular functions. For example, most NADH-binding enzymes have a particular type of folded array - the so-called Rossmann fold which is involved with binding the dinucleotide to accomplish redox reactions (Fersht, 1977). It is usually held that the primary structure of a protein - the sequence of amino-acids - directs the pattern of folding into the secondary structures of alpha-helix and beta-pleat, and the complete folding in three dimensions. Statistical studies have been carried out on the twenty commonly occurring amino-acids to determine which sort of secondary structure they favour and to what extent they contribute to the stabilisation of that structure. The data may now be used for the reverse process - predicting secondary structures from sequences of amino-acids (Chou & Fasman, 1974).

One problem with such prediction is the level of "noise" in the information of a protein sequence. This is due to amino-acids which do not direct folding towards the structure in which they are found - often because of functional restrictions. Tyrosine, for example, reduces the stabilisation energy of an alpha-helix, but it may well be found in a helix involved in DNA-binding. A single such amino-acid will not break up the helix structure, and the structure-forming potential of an amino-acid sequence must be taken over several residues to smooth out

individual residue "noise" from the structure-directing "signal".

Secondly, the pattern of higher folding will have an influence on the secondary structure, one region of secondary structure helping to stabilise another.

Bearing these limitations in mind, an effort was made to predict the three-dimensional structures of filamentous phage capsid proteins. The parameters derived by Chou & Fasman (1974) were used as a basis since they were obtained from a statistical database of proteins without corrections or adjustments. Because the analysis is essentially composed of repetitive calculations based on these parameters, it is eminently suited to processing by computer. The program written for this purpose was called "Protein Fold" and is reproduced in Appendix III. After printing an introduction, shown in Figure 5.1, the program allows sequences to be filed in single-letter code, limited to 70 residues at a time. This arises since the operating system is on a BBC Micro model B, which is limited in the number of 70-element strings and lists it can hold in random access memory (RAM). Protein sequences longer than 70 residues must be cut into overlapping sequences for analysis. This inconvenience was felt to be outweighed by the portability of the program - there are many BBC microcomputers around - and the fact that it can be operated without recourse to a costly time-sharing mainframe computer.

Once an analysis has been commenced, the program first puts up a series of index markers to indicate position within the seventy residue

```

*** PROTEIN FOLD ***      ... by FRANK CSELIK ...

THIS PROGRAM PREDICTS PROTEIN SECONDARY STRUCTURE FROM AMINO ACID SEQUENCES.
THE SEQUENCES MUST BE FILED ON DISC USING THIS PROGRAM.
A DERIVATIVE OF CHOU-FASMAN ANALYSIS IS USED GIVING OUTPUT IN SEVEN LINES :

      : THE FRAME POSITION MARKER, EVERY 10 RESIDUES
DLLAVGGKPM THE AMINO ACID SEQUENCE IN SINGLE LETTER CODE
      ||| ALPHA HELIX INHIBITION | > *
0000-      ALPHA HELIX POTENTIAL 0 > 0 > - ) ASSIGNED AT THE
              > CENTRE OF EACH
VVVV-      BETA SHEET POTENTIAL V > V > - ) PENTA-RESIDUE
      ***| BETA SHEET INHIBITION | > *
-.... +.. POLARITY : HYDROPHOBIC (.), HYDROPHILIC ( ), CHARGED (+) OR (-)
DO YOU WISH TO FILE A SEQUENCE ? N
PLEASE GIVE THE NAME OF THE SEQUENCE TO BE ANALYSED .XFDB

```

```

      :
AEGDDPAKAAFDLSQASATEYIGYAWAMVVVIVGATIGIKLFKKFTSKAS
      |
..- -000000-000000- -000000- -000- - ..
      -----VV-VVVVVVVVVVV-VVVVV----- ..
|***|
..- -..+...- . . . . . +...++ . +.

```

```

DO YOU REQUIRE ANOTHER ANALYSIS ? N
DO YOU WISH TO FILE A SEQUENCE ? N
DO YOU REQUIRE ADVICE ON THE INTERPRETATION OF OUTPUT ? Y

```

ALPHA HELICES AND BETA PLEATS PROPAGATE BI-DIRECTIONALLY FROM INITIATION ZONES (0000) OR (VVVV) ALONG THE POLYPEPTIDE. THEY STOP AT INHIBITION SEQUENCES (||) OR (\*|\*) OR (|\*|) OR (\*\*|) OR (\*\*\*).

P D E ARE STABILIZING AT THE N-TERMINAL END OF AN ALPHA HELIX REGION.  
H K R ARE STABILIZING AT THE C-TERMINAL END OF AN ALPHA HELIX REGION.

THE OUTPUT SHOWS RELATIVE STABILIZATIONS OF THE COMPETING STRUCTURES. THE STRUCTURE OF CONSISTENTLY HIGHER PROBABILITY IN A REGION SHOULD BE ASSIGNED. SOME SEQUENCES - ESPECIALLY OF PHAGE PROTEINS - CAN ADOPT EITHER CONFORMATION. THIS IS SHOWN QUITE CLEARLY IN THE OUTPUT AND IS NOT USUALLY THE RESULT OF AMBIGUITY. RUNS OF (0000) AND (VVVV) SUGGEST ALTERNATIVE STRUCTURES THAT MAY BE STABILIZED ACCORDING TO ENVIRONMENT.

POSSIBLE MEMBRANE INSERTION SEQUENCES ARE ALSO DISPLAYED AS RUNS OF (....) AND ( ) IN THE POLARITY PROFILE WITH NO CHARGES BUT FLANKED BY CHARGED RESIDUES ACTING AS ANCHORS.

```

DO YOU REQUIRE ANOTHER ANALYSIS ? N
DO YOU WISH TO FILE A SEQUENCE ? N
DO YOU REQUIRE ADVICE ON THE INTERPRETATION OF OUTPUT ? N
BYE-BYE

```

Figure 5.1. Sample Output of The Secondary Structure Predicting Program "Protein Fold".

Figure 5.1. Sample Output of The Secondary Structure Predicting  
Program "Protein Fold".

```

*** PROTEIN FOLD ***          ... BY FRANK CSELIK ...

THIS PROGRAM PREDICTS PROTEIN SECONDARY STRUCTURE FROM AMINO ACID SEQUENCES.

```

frame (Fig. 5.1). Below this the single letter sequence is printed, N-terminus to C-terminus. Alpha-helix and beta-pleat potentials are assigned on lines 4 and 5 respectively. There is a hierarchy of symbols in each line to show increasing potential for the structure. This potential is calculated over a penta-residue frame centred on the residue at which the result is reported. Above and below these sequences are the corresponding analyses of significant structural inhibition.

These are the main structure prediction lines, and may be interpreted according to the advice resident in the program (Fig. 5.1). Essentially, secondary structure is regarded as propagating from certain nucleating zones along the sequence until stopped by a corresponding inhibitory zone, or a region more likely to be in the alternative structure is reached. Alpha-helices are shown highlighted in pink, beta-pleats in blue. Zones which have neither of these structures may be in an irregular conformation or in the form of a beta-turn. The latter is essentially a piece of beta-pleat turning back on itself in a hairpin. Below these lines is the polarity profile, which shows up potential membrane-insertion sequences (highlighted in green).

The sequence data on which the predictive analysis was performed were obtained by computer transcription of the viral DNA sequences. The sequence of the fd DNA has been published (Beck *et al.*, 1978), and that of Pf1 has been determined recently in this laboratory (N.J. Short, pers. comm.). Genes 5 and 8 of Pf1 can be recognised from the known protein sequences. These fit on the Pf1 DNA in the same pattern as they

frame (Fig. 5.1). Below this the single letter sequence is printed, N-terminus to C-terminus. Alpha-helix and beta-pleat potentials are assigned on lines 4 and 5 respectively. There is a hierarchy of symbols in each line to show increasing potential for the structure. This potential is calculated over a penta-residue frame centred on the residue at which the result is reported. Above and below these sequences are the corresponding analyses of significant structural inhibition.

These are the main structure prediction lines, and may be interpreted according to the advice resident in the program (Fig. 5.1). Essentially, secondary structure is regarded as propagating from certain nucleating zones along the sequence until stopped by a corresponding inhibitory zone, or a region more likely to be in the alternative structure is reached. Alpha-helices are shown highlighted in pink, beta-pleats in blue. Zones which have neither of these structures may be in an irregular conformation or in the form of a beta-turn. The latter is essentially a piece of beta-pleat turning back on itself in a hairpin. Below these lines is the polarity profile, which shows up potential membrane-insertion sequences (highlighted in green).

The sequence data on which the predictive analysis was performed were obtained by computer transcription of the viral DNA sequences. The sequence of the fd DNA has been published (Beck *et al.*, 1978), and that of Pf1 has been determined recently in this laboratory (N.J. Short, pers. comm.). Genes 5 and 8 of Pf1 can be recognised from the known protein sequences. These fit on the Pf1 DNA in the same pattern as they

do in fd DNA, and the correspondences of the other genes may be inferred. The open reading frames thus assigned for genes 3, 6, 7 and 9 have been shown to give rise to protein components of viral capsids (in the previous chapter), and their secondary structures are to be predicted. This should provide information complementary to the analysis of the DNA sequence (N.J. Short, work in progress).

### 5.2 Phage Proteins of Known Structure

The program was used for a variety of protein-folding predictions, and in particular the predicted structures of fd g5p, fd g8p and Pf1 g8p were studied carefully as reference structures which have been reasonably well characterised. The g5p has been crystallised and the structure is known to very high resolution from X-ray diffraction work (McPherson et al., 1979). The secondary structure prediction of the protein is given in Figure 5.2, together with the structure shown by X-ray diffraction. A very close correspondence is visible. Three shorter regions of beta-pleated structure are found towards the N-terminus of the polypeptide, and the C-terminal half consists of two longer stretches of beta-pleat.

The g8p of fd phages is less well characterised, but the following points are known:

1. In the virion, the molecule is not detectably different in CD from a single rod of alpha-helix;





2. Inserted into the membrane, residues 21 to 39 are inside the membrane and in a beta-pleat conformation. Some random coil is present, but 30% of the protein is still in alpha-helix.

(Ohkawa & Webster, 1981; Chamberlain et al., 1978; Williams & Dunker, 1977; Nozaki et al., 1976; Day, 1966).

The structure prediction shows both possible forms. In this the present program differs from most of its predecessors which only display the single most stable structure - strictly, the most stable structure in a polar environment. Displaying both alpha- and beta- potentials separately allows regions of high stability in both structures to be detected. In the all-helix form, 90% alpha-helicity is predicted (Fig. 5.3). The random structure of the N-terminal five residues has not been noted previously. This could be significant for the mobile charge sequence would lie on the outer surface of the capsid, and be more exposed for solvating the surface of the virion.

The mixed form is predicted to contain 30% alpha-helix, the beta-pleat potential being overriding until Tyr-21. A beta-pleat content of 58% is shown, with 12% random coil. This is in reasonable agreement with the spectroscopic data. Note that a chargeless, non-polar sequence is found from Tyr-21 to Ile-39 (green in Fig. 5.3) which is the zone known to insert into phospholipid membranes (Ohkawa & Webster, 1981; Chamberlain et al., 1978; Woolford & Webster, 1975). Charges on either side of the sequence lock it in position as a membrane-spanning region (Rogers et al., 1980).



In the case of Pf1 g8p, it is known that the capsid form is an interrupted alpha-helix (Makowski *et al.*, 1980). The interruption is shown as a zone of alpha-helix inhibition, Gly-Gly-Tyr, at residues 23-25 in the analysis (Fig. 5.3). An alternative beta-pleated structure is also evidenced, together with a membrane-insertion sequence in the same zone. This shows considerable structural parallels with the fd protein, and suggests that the same type of conformational transition may occur at membrane-insertion in both cases.

It is significant that in both cases the beta-pleat potential towards the C-terminus is a longer and stronger sequence than the alpha-helicity. The relaxed state of the protein must be the one of mixed conformation, not the solid alpha-helix. This is supported by the fact that g8p spontaneously adopts this structure under a variety of circumstances [inserting into phospholipid vesicles (Williams & Dunker, 1977) being solubilised by detergents (Nozaki *et al.*, 1976; Makino *et al.*, 1975) on exposure to hydrophobic solvent surfaces (Griffith *et al.*, 1981)]. The reverse transition is seen only during phage assembly. Thus the mixed alpha/beta conformation is the relaxed state found in hydrophobic environments, while the tense state is the all alpha-helical conformer found in capsids. The transition from tense to relaxed state is triggered by a hydrophobic environment, while that from relaxed to tense has not yet been described in any detail.

### 5.3 The Structures of The g3ps

The g3p is known to be located in the membrane of an infected host cell (Jazwinski et al., 1973), where it awaits addition to the last end of the virion to emerge across the membrane (Lopez & Webster, 1984; Armstrong et al., 1981). The membrane spanning region is towards the C-terminus of the protein (Boeke & Model, 1982; Armstrong et al., 1981)). Such sequences have been analysed in various membrane proteins, and show a region of charge free, largely non-polar amino-acids, which stably interact with the non-polar lipids of the membrane. Flanking this sequence on either side are charged residues which may interact with the polar ends of phospholipid molecules but are thought primarily to provide anchors against the hydrophobic region slipping through the membrane. Dragging these charges through the hydrophobic membrane interior would demand a high energy of activation (Rogers et al., 1980).

Predicted membrane anchor sequences have been found on the C-terminal segments of g3p from both fd and Pf1 (green in Fig. 5.4). The corresponding Pf1 sequence shows a particularly hydrophobic run of residues. Both sequences show considerable alpha-helix and beta-sheet potential (pink and blue respectively). Thus they may well follow the pattern of the g8p in having a membrane inserted beta-pleated conformer, and a capsid alpha-helical form.

The N-terminus of the g3p was also analysed, with a view to finding the signal peptidase clipping site (Fig 5.5). As shown in the previous

```

# # # #
YVFGAGKPYEFSIDCDKINLFRGVFAFLLYVATFMYVFSTFANILRNKDS
* |
--- -- --00000--0- - --0-
VVV- - V-----V-VVVVVVVVVVVVVVVVVVVVVV--VV-----
**| **|
. . +. -. . -+. . +. . . . . . . . . . . . . . . . +. +.

```

```

# # # #
GGGQLPCESMTIWGKSYSCLISDYAGQLSSLRVALLLMAALIAALILLKD
||| | ***
--- -- -- -- --00000000000000000000..
--- --VV-----VVVV-----VVVVVVVVVVVVVVVVVVVV..
*** || **| ||
. . - . . + . . - . . . +. . . . . . . . . . . . . . . . +.

```

Figure 5.4. Membrane anchor sequences found in the C-terminal regions of the g3p of fd (above) and Pf1 (below). In both cases a long run of charge-free residues (green) is flanked by charged ones. Note the extremely hydrophobic run of residues (.) in the case of Pf1 g8p.

```

      #           #           #           #
MKKLLFAIPLVVPFYSHSAETVESCLAKPHTENSFTNVWKDDKTLDRYAN
      |           |           |           |
..00000000- - 00000-00--0-- - - - - - - - - - -
..V--VVVVVVVVV - - - - - - - - - - -V-- - - - - -
      |||      ||      *|| |||      *****
..++ ..... + - - - - - ++ - - - - - + - - - - -

```

```

      #           #           #           #
MSIKTLISVLRVTLLTACLLPSLFFVRSAGPYIWEVVMYSSSSGSSTPA
      |           |           |           |
.. - - - - -0- - - - -00000 - - -00- - -0- - - -000 - -0
..V-V-VVVVVVVVVVVVVV- - -VVV-V- - - - - -VVVVVVV-
      ||           ||           ||*||
.. ++ .. ++ ..... ++ ..... ++ ..... ++ .....

```

Figure 5.5. N-terminal sequences of the g3p from fd (above) and Pf1 (below). Each shows a run of chargeless residues (green) flanked by charges. In both cases the sequence at the C-terminal end of this is (+)residue-Ser-Ala, the Ala forming the N-terminus of processed fd g3p. A similar cleavage site is indicated in the Pf1 protein (arrowed).

chapter, Pf1 g3p has a higher electrophoretic mobility than the sequence of its gene would suggest (Short, pers. comm.). This may well be due to the cleavage of an N-terminal extension from the nascent polypeptide, as is the case for fd g3p. In the case of fd g3p a very clear linear structure is seen: anchor charges - membrane insertion sequence - (+) charge - polar residue - non-polar residue, with the cut site located between the polar and non-polar residues. A parallel pattern is found in the Pf1 g3p sequence. Indeed the sequence around the known cut site of fd g3p is His<sup>+</sup> - Ser - Ala, and the sequence Arg<sup>+</sup> - Ser - Ala is found in the corresponding location on the Pf1 sequence. Thus it is very likely that this is the cut site of signal peptidase, producing a Pf1 g3p with Ala-29 as the N-terminus.

#### 5.4 The Structures of The g6ps

The g6p of the Ff phages is located at the g3p-end of the capsid (Grant *et al.*, 1981). Some amber mutants in gene 6 show a normal phage morphology but have relatively unstable capsids which are readily disaggregated (Marvin & Hohn, 1969). How the protein locks the capsid structure is unknown, but some peculiar features emerge from structure prediction. The fd g6p contains the ambivalent alpha-helix / beta-pleat potentials found in other phage proteins. It is extremely hydrophobic, the most hydrophobic product of the genome (Appendix II). Three runs of non-polar residues are seen, any or all of which may insert into the membrane (Fig. 5.6). Yet the molecule also bears 8 (+) charges - all but one due to Lys - and 5 (-) charges - all due to Asp. The locking of the

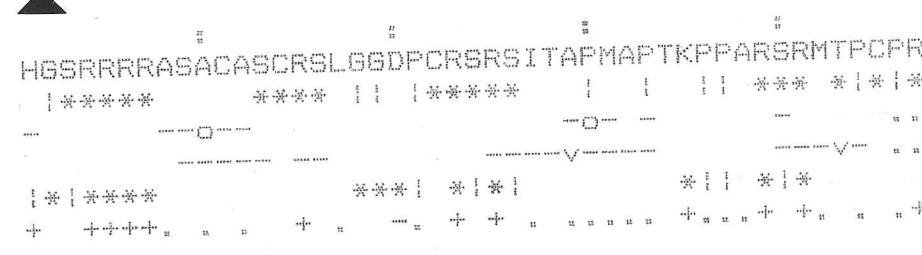
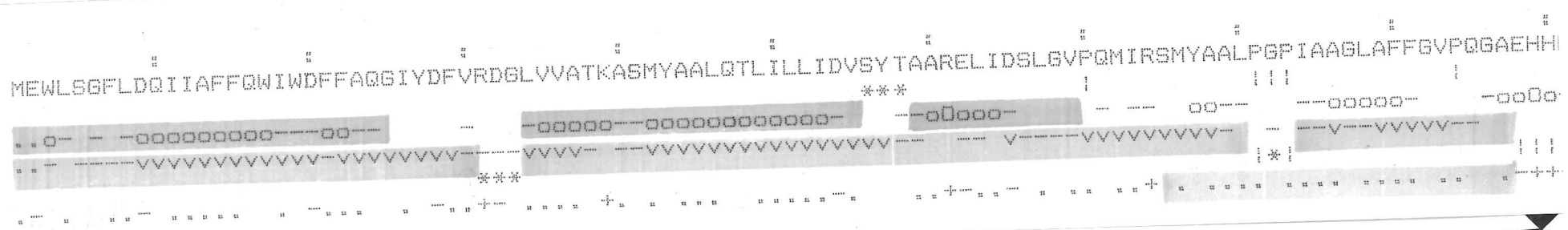
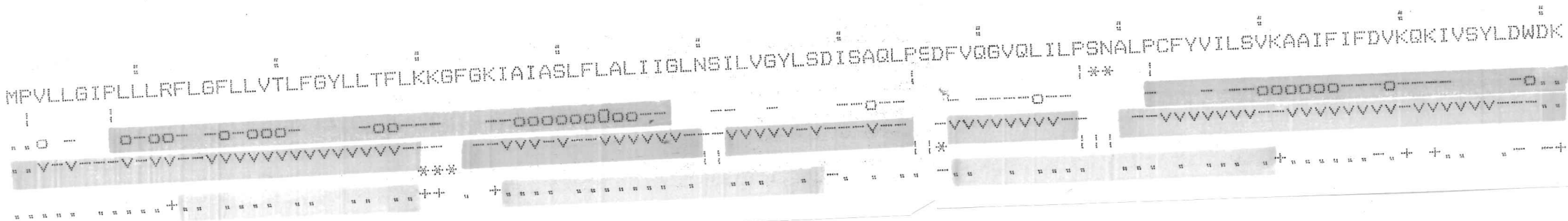


Figure 5.6. The predicted secondary structures of the g6p of fd (above) and Pf1 (below). Both show regions of high alpha-helix (pink) and beta-pleat (blue) potential. The fd g6p exhibits three potential membrane-insertion sequences (green), but that of Pf1 only one. A notable feature of the prediction for Pf1 g6p is the unstructured nature of the C-terminal 60 residues, and the cluster of charges in this region. There appear to be no obvious similarities in the predicted secondary structures.



Figure 5.6. The predicted secondary structures of the g6p of fd (above) and Pf1 (below). Both show regions of high alpha-helix (pink) and beta-pleat (blue) potential. The fd g6p exhibits three potential membrane-insertion sequences (green), but that of Pf1 only one. A notable feature of the prediction for Pf1 g6p is the unstructured nature of the C-terminal 60 residues, and the cluster of charges in this region. There appear to be no obvious similarities in the predicted secondary structures.

phage capsid structure may be due to powerful intersubunit interactions mediated by the hydrophobic sequences. Such additional binding is necessary, for the linear (helical) symmetry of the capsid is broken at the ends, and capsid structure is correspondingly weakened.

The corresponding protein from Pf1 also contains ambivalent sequences but has many features that contrast with the fd g6p (Fig. 5.6). Although still the most hydrophobic protein of its genome, there is only one significant potential membrane insertion sequence. The molecule is far more highly charged, having 18 (+) charges - predominantly due to Arg (not Lys) - and 10 (-) charges mainly due to Asp (as in fd g6p). Much of the charge density is located within 50 residues of the C-terminus. This area is predicted to have very weak secondary structure with much structural inhibition. A beta-sheet might be able to form if there were good enough lateral hydrogen bonding to hold it in place. Since this region has a net charge of 12 (+), it may interact either with the DNA, like the C-terminal part of the g8p molecules, or with the negatively charged outside of the capsid.

### 5.5 The Pilot Protein Theory

The importance of minor capsid proteins in assembly was postulated some time ago but only g3p was then known as a minor component of filamentous viral capsids. At that time it was proposed that the g3p not only guides infection of the appropriate host, but also pilots the DNA through various stages of its replication, finally initiating the

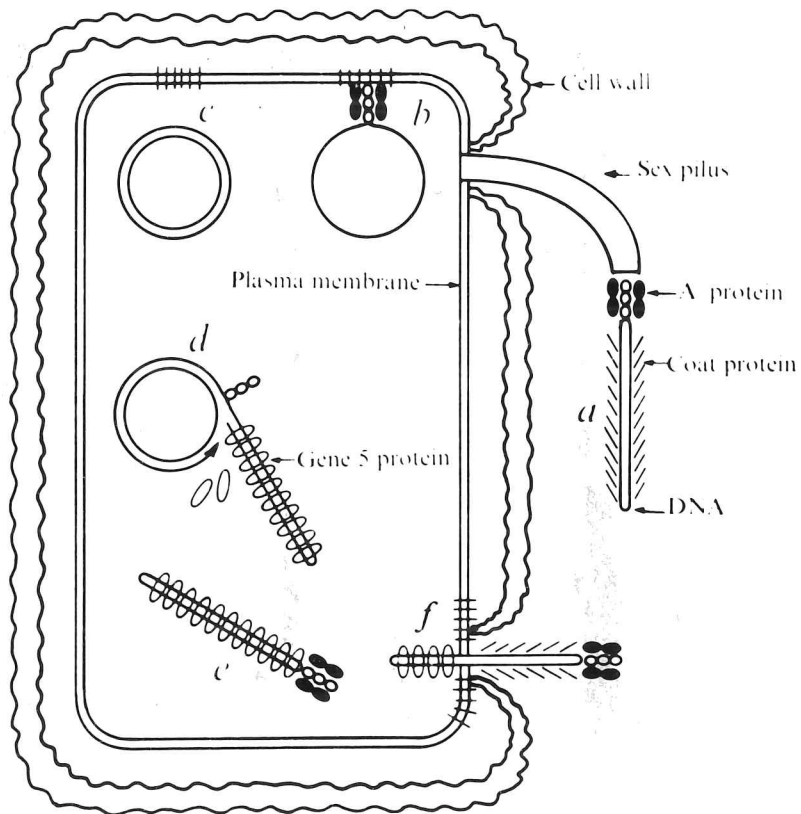


Figure 5.7. Diagram of the life-cycle of fd phage. Pilot proteins (black ellipses) are shown docking the viral DNA to the membrane (b), at one end of the intracellular g5p.DNA complex (e), and directing the assembly process (f). Mature virions are shown being extruded from the cell, pilot protein end first.

assembly of capsid proteins around the DNA to form complete virions (Zlochevski *et al.*, 1977; Jazwinski *et al.*, 1973). The polyfunctional protein was thought to have several separated active sites, each related to a separate piloting function in the life-cycle (Kornberg, 1982; Kornberg, 1974). This view must be revised, first in the light of the finding that filamentous phage capsids contain four sets of minor protein components, two at either end of the capsid. Any piloting functions could therefore be divided between these proteins. Secondly, more recent studies show that the g3p-end of a virion is the last part of the phage to be released (Lopez & Webster, 1984; Armstrong *et al.*, 1981). The initiation of assembly must therefore be related to the g7p.g9p if minor coat proteins are involved, since these are at the opposite end from g3p (Fig. 5.7). It might be expected that, like the g3p, the initiator proteins would show considerable homology of structure, owing to conservancy in evolution. This is especially so because of their small size, for single amino-acid substitutions amount to large proportionate changes of the overall structure.

#### 5.6 Possible Pilot Proteins - g9p and g7p

The g9ps of the two phage classes show considerable disparity in secondary structure predictions, not least because of their size differences: the fd protein is 32 residues long whereas that of Pf1 is 83 residues in length. The fd protein is predicted to have a predominantly beta-pleated structure, with a break around residue 19, probably as a beta-turn (Fig. 5.8). A rather unsatisfactory membrane insertion sequence might be present from the N-terminus to Arg-17, but would lack a charged membrane anchor sequence at the N-terminal end.



In contrast, the Pf1 protein appears to have a clear membrane insertion sequence with 3 (+) and 1 (-) charges as anchors at either end (Fig. 5.9). The predicted structure is predominantly beta-pleat, but there is a large alpha-helical zone from Pro-17 to Ile-34, and possibly one from Pro-54 to the C-terminus. The considerable dissimilarity in predicted structure between the fd and Pf1 g9p indicates that either the proteins are adapted to different functions in the phages, or that they are not stringently adapted to a common function.

Nothing could be further from the truth when considering the structures of g7ps. Both the fd and Pf1 proteins are roughly the same length (33 and 30 residues, respectively). Both reveal considerable helix and pleat potential (Fig. 5.9). Each has a membrane insertion sequence bounded by charged residues, positive (exclusively Arg) on one side and negative (mainly Asp) on the other. This high degree of structural similarity would be consistent with these proteins being specifically adapted to an important function.

The g7p is, therefore, a good candidate for a morphogenetic signal initiating assembly. In keeping with the original concept of a pilot protein (Kornberg, 1974) the g7p may be bound to the DNA during replication, direct the assembly of the g5p.DNA complex, and then initiate the final assembly of the viral capsid. In this way the g7p may be a pilot protein for the assembly process. This would be consistent with the position of the protein at the end of the capsid first extruded (Lopez & Webster, 1984; Armstrong *et al.*, 1981), which also contains a

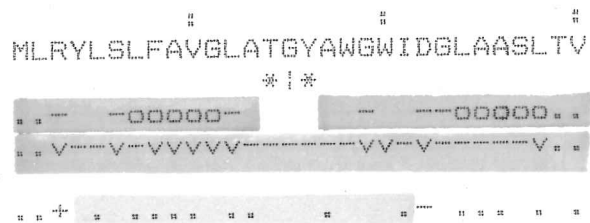
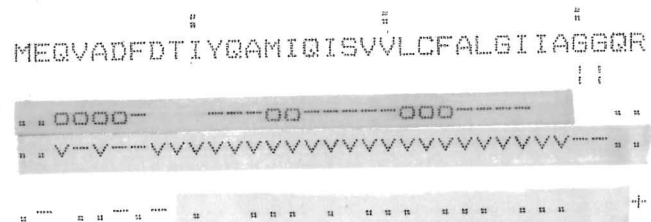


Figure 5.9. The predicted secondary structures of the g7p of fd (above) and Pf1 (below). Strong similarities are visible: each protein is of about the same size, having strong potential for alpha-helix (pink) and beta-pleat (blue), and a membrane insertion sequence (green) bounded by charged residues.

large DNA hairpin. That hairpin (A) has been shown by deletion studies to be necessary for the assembly of phages (Dotto et al., 1981b), and would form an easily recognised structural feature in the DNA to mark the origin of assembly. It is thus possible that the initiation of assembly depends on an interaction between the A-hairpin and the capsid proteins, mediated by g7p. The next chapter describes an analysis of the g5p.DNA complexes of phages fd and Pf1 to investigate the stage of virus assembly at which the minor coat proteins are added to the eventual virion.



## CHAPTER 6

### THE PROTEINS OF G5P.DNA COMPLEXES

### 6.1 The g5p.DNA Complexes

The essential feature of the production of mature virions is the process of linear assembly, with initiation, extension and termination phases triggered by the properties of the assembling components. The bacterial membrane divides this system into two compartments: the cytoplasm and the external medium. In the cytoplasm the DNA committed to assembly is associated mainly with g5p; in the medium it is associated mainly with g8p. At assembly, the DNA sheds g5p and is surrounded by g8p in a smooth and continual process.

It had not been determined whether the minor protein components of the capsid are already associated with the DNA in the g5p.DNA complex, or waiting in the membrane with the g8p. Following the original version of the pilot protein concept, one or more of the proteins would be associated with the DNA during or soon after replication (Kornberg, 1974). Thus they could be present on the intracellular precursor of phage, the g5p.DNA complex. The g7p.g9p could be associated with the DNA hairpin, and the g6p with the DNA at the other end of the complex. The pilot proteins would signal the initiation of assembly, followed by elongation of the particle as g8p subunits from the membrane replaced g5p subunits on the growing capsid. The g6p could terminate this process, and lock the capsid structure as discussed previously (section 5.4). It is notable in this context that some mutants in g3p and g6p cause viruses to be synthesised with over-long capsids, indicating some failure in the termination of assembly (Marvin & Hohn, 1969; Salivar et

al., 1967). The g3p is known to be harboured in the cell membrane and to be added to the last end of the phage to emerge from the infected cell (section 5.4).

The extreme alternative model would regard these proteins as membrane bound. The analyses presented in the previous chapter show that the proteins do have potential membrane insertion sequences. The g7p.g9p could thus be envisaged as a membrane-bound receptor site for the exposed DNA hairpin of the g5p.DNA complex. This is likely to be exposed, for g5p binds only single-stranded DNA and the hairpin is, of course, double-stranded (assuming it to have formed). The DNA would lock into the receptor to trigger assembly of the g8p into capsid, which would propagate until all the DNA was covered. The g3p.g6p would then seal the end of the capsid. To the demerit of this hypothesis, it is difficult to see how the g3p.g6p could attach to the virion at the correct point in assembly, for there do not seem to be any specific DNA-protein interactions at this end of the capsid (Webster et al., 1981). Yet they would presumably have to recognise the end of the virion as it traversed the membrane.

Between these extremes, models may be proposed which involve fewer pilot proteins. For example the DNA hairpin-A may be complexed with g9p only, providing the initiation signal, with g6p providing the terminator. Alternatively both g7p and g9p may be present in the g5p.DNA complex, but terminator proteins may be completely absent. There is a great diversity of these intermediate models which it would be trivial to enumerate.

## 6.2 Purification of the fd Complex

In order to investigate the g5p.DNA complex of fd phage, that nucleoprotein must first be isolated. Since the subjects of the study are minor protein components, it must be very pure indeed. Methods of procuring the g5p.DNA complex at the high purity required were thus investigated briefly.

A preliminary experiment was performed to ensure that the g5p from g5p.DNA complexes could be located on polyacrylamide gels. E.coli K38 cells were grown in 10 ml cultures of 2xTY medium, and infected with phages at moi 100 in mid-log phase growth. Only 3 h of culturing were allowed after infection, as the cells had reached maximum density by this time and further culturing would have been pointless. The cells were pelleted by centrifugation at 10,000 x g for 10 min at 4°C and resuspended in 1 ml of 8 M urea, 2% SDS, 10 mM Tris.HCl, pH 7.5, 1 mM EDTA. Incubation at 37°C for 20 min gave a clear yellow solution from which samples were taken for electrophoretic analysis. A sample of total protein from uninfected cells was prepared in parallel. The gel showed two phage-specific bands (Fig. 6.1). The lower band migrated with slightly lower mobility than mature g8p, and was assumed to be the slightly larger pro-coat (nascent g8p, Wickner, 1983; Date & Wickner, 1981; Ito et al., 1980). The upper band was of approximately the mobility expected of g5p ( $M_r$  9,690).

(a)



(b)



Figure 6.1. Infection-specific proteins in E.coli K38 cell lysates. Track (a) contains about 0.1 mg total protein from a lysate of E.coli K38 cells. The lysate in (b) was from the same strain of cells infected with phage fd. Two infection-specific bands are visible (arrowed). The mobility of the upper one indicates that it is g5p, and that of the lower one that it is the nascent g8p, the pro-coat protein.

In order to investigate the purification of the complex, a 20 ml culture of E.coli K38 was prepared in MTPA medium, infected with phages and harvested after 3 h by centrifugation at 10,000 x g for 10 min at 4°C (section 2.3). The cells were resuspended in NT-buffer and resedimented. This washing step removed phages which were trapped in the pellet or adhering to the bacterial cell walls. Such a step is of crucial importance since both phages and complexes are nucleoproteins of similar size, shape and density: thus they would be very difficult to separate at a later stage.

The final resuspension of cells was found to be best in NET-buffer to an  $A_{680}$  of 1.0. A significantly different optical density of cells gave a very poor overall yield of complexes. The usual method was to resuspend in 1/4 the culture volume of NET-buffer, determine  $A_{680}$ , and dilute as necessary. To this suspension were added 1/9th volume of 11 mg/ml lysozyme and 1/100th volume of 0.2% PMSF in ethanol. The former lysed the cells in the hypo-osmotic buffer, the latter inhibited cell proteinases which might have degraded the complexes. After being stirred for 1 hour at room temperature, the suspension was cleared by centrifugation at 25,000 x g for 30 min at 4°C, and the supernatant subjected to centrifugation at 100,000 x g for 5 h at 4°C. This sedimented complexes with a yield of approximately 2.5 mg/l cell culture. The pellet showed considerable impurity in Simons electrophoresis and silver-stain.

Further purification of the complex was essential before any radiolabelling experiments could be done; the bands of impurity would obscure any minor protein bands present on a fluorogram. A 5-20% sucrose density gradient was prepared, and complex purified by velocity sedimentation in a Beckman SW 50.1 rotor at 32,000rpm for 1 hour at 20°C. Banding of the material was diffuse as shown by negative displacement of the gradient through a recording spectrophotometer. Peak fractions were dialysed against glass distilled water for 24 h and freeze-dried, then analysed by electrophoresis and silver-stain (Fig. 6.2). Samples from different parts of the peak showed different proportions of the same impurities, though the major band in each case was g5p.

A better method for purifying nucleoproteins is isopycnic density gradient centrifugation. The nucleoprotein should band at a greater buoyant density than impurity proteins, and so should be well separated by this technique. Unfortunately, the g5p.DNA complex is very salt-labile, and would not survive the necessary concentration of caesium chloride involved. A dense glucose derivative, 2-(3-Acetimido-5-N-methylacetimido-2,4,6 tri-iodo-benzamido)-2-deoxy-D-glucose (Metrizamide), is available however, and this can form density gradients in an analogous fashion. It should not dissociate the complex, any more than sucrose does. There is a risk that at elevated temperatures the Metrizamide could act as an iodinating agent, so all manipulations were carried out at 4°C.

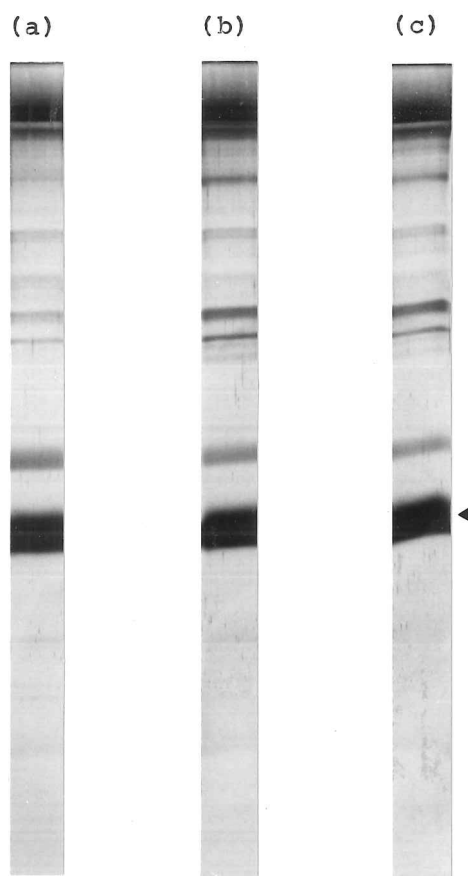


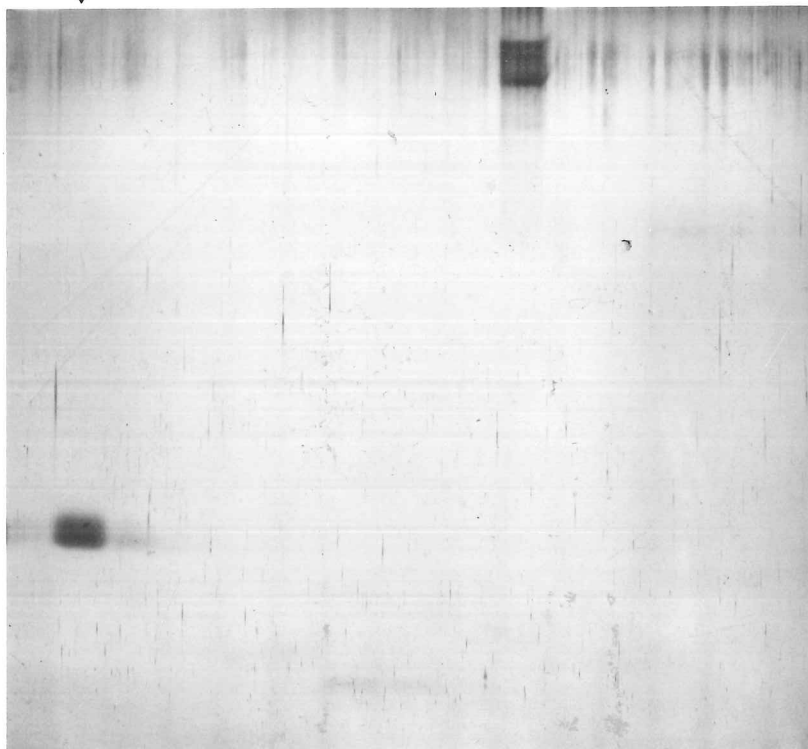
Figure 6.2. Attempted purification of g5p.DNA complexes of fd by sucrose density gradient centrifugation, as monitored by electrophoresis and silver-staining. Sedimentation of the crude product of the purification procedure (section 2.3) through a 5-20% sucrose density gradient failed to separate the complexes from the contaminants. Tracks (a), (b) and (c) are from the leading, central and trailing portions of the peak in  $A_{260}$  of the gradient (g5p indicated). Little purification is achieved in any of them.



A sample of crude g5p.DNA complexes was dissolved in TE-buffer, which was rendered 35% in Metrizamide. This was centrifuged in a Beckman SW 50.1 rotor at 45,000 rpm for 64 h at 4°C. A solution of 5 M NaCl, 0.01% bromophenol blue was found to give the correct density for the counter balance tube. Fractions were collected by negative displacement from the bottom of the tube, then analysed by Simons polyacrylamide gel electrophoresis and silver-staining (Fig. 6.3). The complexes were found to form a narrow band (2 fractions of 18) near the bottom of the gradient. This product was pure, and was removed from the Metrizamide solution by tenfold dilution in TE-buffer and ultracentrifugation at 100,000 x g for 5 h at 20°C. This method is preferred to dialysis on grounds of speed and convenience.

### 6.3 Minor Proteins of the fd Complex

Since highly pure complexes could be isolated, it was feasible to attempt radiolabelling with specific amino-acids to enable minor protein components to be detected in these samples. This was performed by the system detailed in Chapter 4, wherein the development and application of the system was described. [<sup>3</sup>H]-arginine was used as radiolabel since, although arginine is present in fd g5p (Table 4.1), the electrophoretic mobility of g5p is quite different from the mobilities of the phage minor coat proteins (except g6p).



← Sedimentation

Figure 6.3. Purification of g5p.DNA complexes of fd by Metrizamide equilibrium density gradient centrifugation. A self-forming gradient of 35% Metrizamide was loaded with impure complexes and centrifuged, then cut into fractions and analysed by polyacrylamide gel electrophoresis and silver-staining (section 2.4). Sedimentation, as portrayed in the gel, was from right to left. As a nucleoprotein, the g5p.DNA complex (arrowed) bands at higher density (lower) in the gradient than the protein contaminants, and is thus separated from them.

A culture of 10 ml E.coli K38 in MTPA(-Arg) was prepared, and infected with phages, moi 100, at  $A_{680}$  0.16. At the same time, 1 mCi of [ $^3$ H]-arginine monohydrochloride (at 28.8 Ci/mmol) was added. The culture was continued for 3 h, then the cells were harvested, washed and lysed (section 2.3.3). The g5p.DNA complexes were precipitated twice from the cleared lysate, then subjected to isopycnic density gradient centrifugation in 35% Metrizamide as described above. The gradient was cut into 10 fractions, but only a small proportion of each was available for analysis - the rest being required as product. Absorbance in the UV range was not a useful method for locating the small amount of material on the gradient, as the absorbance of Metrizamide swamps that of the protein -  $A_{239}^{1\text{ cm}, 1\text{ M}}$  being 54,000. Detection of protein by electrophoresis and silver-stain was feasible, but agarose gel electrophoresis - detecting the DNA of the complex - would be as sensitive and faster. Furthermore, the material was thinly dispersed in a concentrated sugar solution, so it was not suitable for loading onto protein gels directly.

Accordingly, 5% of each sample was subjected to electrophoresis in a 1% agarose gel (section 2.7). Samples of each fraction were taken for scintillation counting, so that the distribution of radioactivity on the gradient could be compared with that of the DNA. As Figure 6.4 shows, there was considerable radioactivity not associated with the DNA, which was left behind when the DNA-containing fraction was selected. Considerable purification has thus been achieved. This fraction was diluted tenfold with TE-buffer and centrifuged at  $100,000 \times g$  for 5 h at  $20^\circ\text{C}$  to pellet the complexes, which were then resuspended in 0.05 ml TE-buffer.

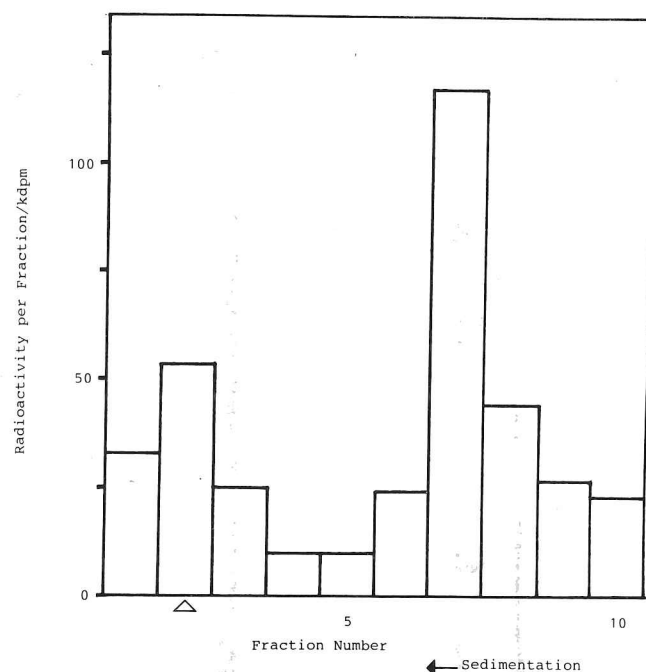


Figure 6.4. Radioactive profile of an equilibrium density gradient in Metrizamide of impure g5p.DNA complex from fd-infected cells. Sedimentation was from right to left. The fraction containing the viral DNA is indicated with an arrowhead. Separation of nucleoproteins from the larger protein peak is less clear than in a corresponding caesium chloride step with phage particles (Fig. 4.2).

The product was divided into two fractions, each of which were denatured with loading buffer and loaded onto Simons polyacrylamide gels. Electrophoresis was performed as standard, followed by PPO impregnation, drying down and fluorography (section 2.4). A reference track of phages was obtained by preparing the radiolabelled virions from the supernatant over the cells, according to the protocol in section 2.4 and 4.2. This showed up minor protein components on the fluorographs to allow comparison with the g5p.DNA complex tracks. The minor proteins of phage capsids were observed as radiolabelled bands, as seen in the earlier experiments (Fig. 6.5). The complex track, however, showed no corresponding bands at the g7p, g9p or g3p positions.

Unfortunately, g6p almost comigrates with the g5p and may thus have been masked. The experiment was therefore repeated by radiolabelling with [ $^3\text{H}$ ]-tryptophan, which should label g6p but not g5p (Table 4.1). The culturing and purification proceeded as described above, except that radiolabel was added as 1 mCi of [ $^3\text{H}$ ]-tryptophan (at 3.4 Ci/mmol). Note that g7p would not be labelled in this system - there is no amino-acid present in g3p, g6p, g7p, and g9p but not in g5p. Radiolabelled phages were obtained from this preparation, but no significant radioactivity was detectable in the complexes isolated. The complexes showed both g5p and DNA in electrophoretic analyses, banding at the correct nucleoprotein density in the Metrizamide gradient. No evidence of other minor protein components was found.

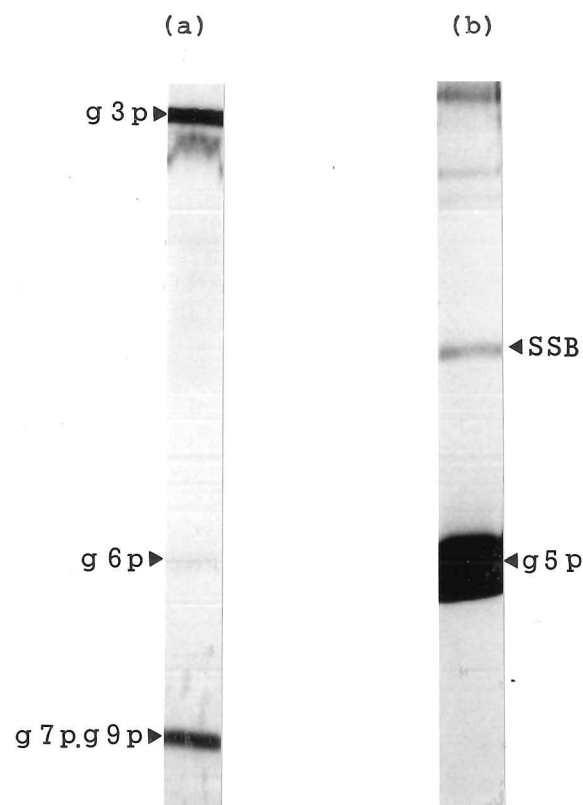


Figure 6.5. Fluorographs of selectively radiolabelled phage fd and its precursor g5p.DNA complex. Track (a) contains phages isolated from a culture of *E.coli* K38 grown in MTPA(-Arg) medium with [ $^3\text{H}$ ]-arginine. The g3p, g6p, and g7p.g9p are visible. Track (b) contains g5p.DNA complexes isolated from a similar culture, and shows no corresponding bands. The major band is the g5p itself, and a minor component is thought to be SSB, copurifying with g5p.DNA complex as the SSB.DNA pre-initiation complex. The minor proteins of the fd capsid are not present in the g5p.DNA complex.

Thus it may be concluded that the g5p.DNA complex of phage fd does not contain any of the minor protein components found in the viral capsid. Not only were the methods used sensitive enough to detect the minor proteins of the capsid, but the g5p.DNA complexes were intact nucleoproteins (shown by the buoyant density in Metrizamide), with the correct DNA (migration in agarose gels) and the expected major protein component (g5p, seen in the fluorographs). The finding that the minor proteins are absent from these complexes seems well founded and appears to rule out any role of the minor proteins as pilot proteins in DNA replication or assembly. One whole class of models described above can therefore be discarded.

An interesting feature of the fluorogram of radiolabelled fd g5p.DNA complexes was the presence of a labelled band at  $M_r$  15,000 which could not be that of any minor coat protein. The most likely explanation is that it is due to SSB protein, of  $M_r$  14,500, which might be expected to bind the single-stranded virion DNA. This would cause it to copurify with g5p.DNA complex as a contaminating SSB.DNA nucleoprotein: the pre-initiation complex. These two nucleoproteins - SSB.DNA and g5p.DNA - are at the main division point of the DNA replicative cycle. A single-stranded DNA produced in replication can either form SSB.DNA complex, and be directed towards replication, or form g5p.DNA complex, and be directed towards assembly (Schaller, 1978). In this, the host cell SSB and the phage coded g5p are acting as regulators and the ratio of SSB to g5p determines the net course of DNA flow. In the absence of g5p at infection, SSB causes RF replication to run exponentially. As g5p

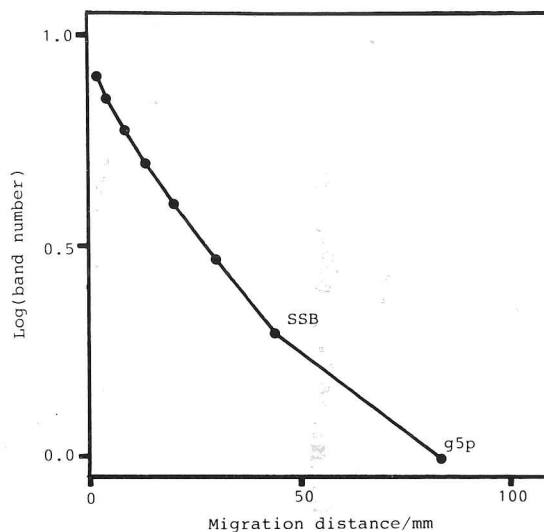


Figure 6.6. Anomalous migration of g5p in polyacrylamide gels. The gel track shows a series of bands at successively lower mobilities behind g5p. The migration distance, plotted against the logarithm of the band number (taking g5p itself as 1) gives a reasonably flat curve. This would normally be interpreted to indicate the formation of a series of multimers of the fundamental protein species, though why g5p should migrate thus is not known.



accumulates, it begins to divert (+) stands towards assembly. This equilibrium also keeps the RF population stable through cell division, for as the cell grows g5p is diluted out and more DNA can pass back into RF synthesis. Both types of protein complex are therefore present in the virus-infected cell and their copurification is a distinct possibility.

Some gels of denatured g5p.DNA complexes showed a whole series of bands at lower mobilities than the g5p. The spacing of these bands was too regular for them to be due to impurities of random relative molecular masses. Plotting  $\log(\text{band number})$  against migration distance (Fig. 6.6) showed that they form a classical "ladder": a series of bands with a constant difference in apparent  $\frac{M}{r}$ . This is usually interpreted as the formation of dimers, trimers, tetramers and higher, of a protein species present in the track. In this case, the unit size is that of g5p. How this should come to form stable multimers in the denaturing conditions employed is not clear. It is not an unusually hydrophobic protein (like g8p) and should not be difficult to disaggregate. The cause may be a direct chemical crosslinking induced by Metrizamide, during the isopycnic density gradient step.

Experimental quirks aside, the g5p.DNA complex of fd phage has been shown to lack the g3p, g6p, g7p and g9p which might have been piloting the DNA to assembly. Indeed, it is apparent that the g5p itself must be responsible for this function, sequestering the DNA from the replicative pool. This, in turn, raises two questions: Do the Class II phages exhibit the same lack of dependence on minor proteins for piloting the DNA to assembly? If the protein does not act as the trigger for the assembly process, what does?

#### 6.4 The Pf1 g5p.DNA Complex

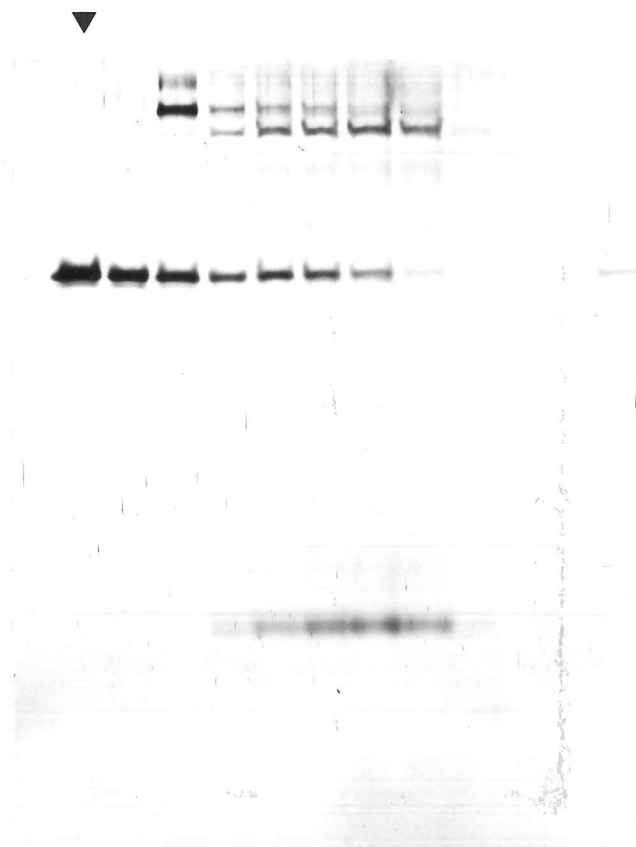
Considerable parallels have been shown between the Class I and Class II filamentous bacteriophages. Both assembled capsids contain a complement of minor protein components coded by genes 3, 6, 7 and 9. It is known that both viruses are assembled from filamentous g5p.DNA complexes (Kneale & Marvin, 1982; Pretorius *et al.*, 1975; Pratt *et al.*, 1974). The minor proteins have been shown to be absent from the g5p.DNA complexes of the Class I phage fd. It is of considerable interest to determine whether this is also true of the g5p.DNA complexes of the Class II phage Pf1, or whether these phages show a different approach to assembly, reflecting their different assembled morphology.

The specific radiolabelling approach was used to render the minor proteins of phage Pf1 detectable. A culture of *P.aeruginosum* K was prepared in 10 ml MTPA(-Trp) and infected with phages moi 100 at  $A_{680}$  0.16. At the same time, 1 mCi [ $^3\text{H}$ ]-tryptophan (at 3.4 Ci/mmol) was added to the medium. The culture was propagated for 5 h before harvesting the cells to allow stationary phase to be reached. Washing and lysis of the cells was as described for the fd g5p.DNA complex purification. The only difference was the higher background of radioactivity observed on the Metrizamide gradient, and the higher total yield of radioactivity in the final product. Analysis by Simons electrophoresis and fluorography showed that this was because the complexes were still contaminated extensively with other proteins of high relative molecular mass. This was despite the fact that only one fraction, containing phage DNA, had

been selected from the Metrizamide gradient. The isolation protocol therefore had to be redesigned for better purification of Pf1 g5p.DNA complexes.

This was broken down into two sub-problems: increasing the yield of complex, both to increase its proportion in the product and to allow further purification steps to be added; and increasing the selectivity of the purification for g5p.DNA complexes. A major problem relating to the former was the considerable loss of intact cells during the washing phase. P.aeruginosum cells were quite difficult to resuspend, even if sucked up and down in a Pasteur pipette. Full resuspension seemed to be accompanied by considerable mechanical damage to cells, for the size of the pellet decreased markedly on each round of washing. To minimise this source of loss, one round of washing was eliminated. The period of lysis with lysozyme was also extended to 2 h, to ensure that all the cells had released their contents.

A sample of g5p.DNA complexes was prepared in this way, from a 100 ml culture of Pf1-infected P.aeruginosum K in MTPA without radiolabelling, and subjected to Metrizamide density gradient centrifugation. The fractions were collected and samples analysed by Simons polyacrylamide gel electrophoresis and silver-staining. The result is shown in Figure 6.7, and the cause of contamination is clearly visible. Some proteins of high  $M_r$  were banding at an abnormally dense level in the gradient. This suggests that either they were part of some other intracellular nucleoprotein complexes - such as ribosomes - or contained heavy metal



← Sedimentation

Figure 6.7. Attempted purification of g5p.DNA complexes of Pf1 by Metrizamide equilibrium density gradient centrifugation, analysed by polyacrylamide gel electrophoresis. Sedimentation is from right to left. The region in which g5p.DNA complexes attain their buoyant density (arrowed) overlaps with that in which contaminants settle. Thus only a limited purification is achieved.

ions as prosthetic groups, increasing their density. The peak was not coincident with the g5p.DNA complexes, so the proteins are not part of those nucleoproteins, but there was considerable overlap between the peaks. Thus the peak fractions of the complexes contained some of this dense protein contaminant.

It was essential to remove the contaminant before the Metrizamide density gradient centrifugation, to reduce this overlap. One separation method not used in the protocol thus far was precipitation. This has been used as a matter of necessity in purifying large quantities of the complexes, since it is much more convenient to precipitate from large volumes of cell lysates than to ultracentrifuge them (Kneale & Marvin, 1982). A further radiolabelled culture was prepared. Labelling was with 1 mCi [ $^3\text{H}$ ]-tryptophan (at 3.4 Ci/mmol) in 10 ml MTPA(-Trp), and the culture was grown up as normal (section 2.3). The cleared cell lysate was made 5% PEG 6000, 0.5 M NaCl by the addition of these compounds as solids. A precipitate was allowed to form overnight, and was collected by centrifugation at 15,000 x g for 90 min at 4°C. Following the precedent of the phage purification methodology (Chapter 3), a second PEG purification step was tried in the hope that additional purification would be achieved. The second precipitate was collected by centrifugation and resuspended in 3 ml ice-cold TE-buffer, in which were dissolved 3 g Metrizamide to make 5 ml total. The solution was subjected to ultracentrifugation in a Beckman SW 50.1 rotor at 45,000 rpm for 64 h at 4°C, then fractionated. The fractions were analysed by agarose gel electrophoresis to localise the DNA, the relevant fractions being

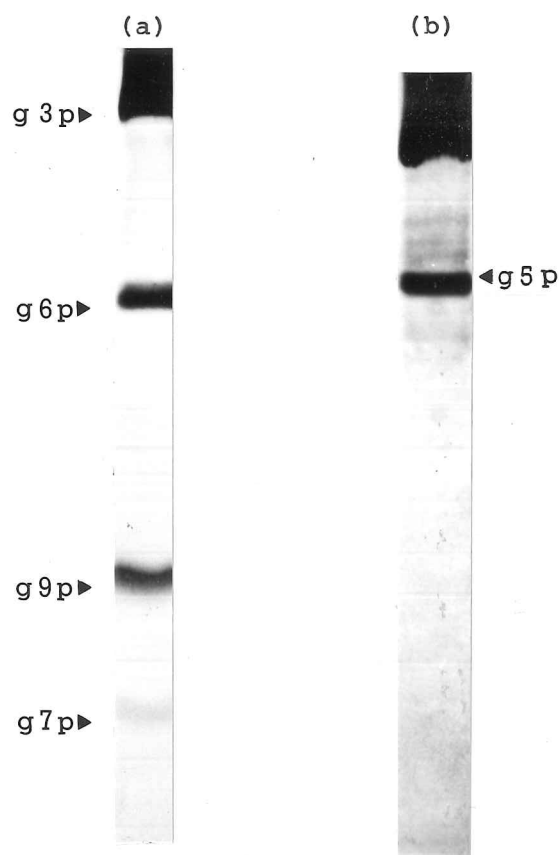


Figure 6.8. Fluorographs of selectively radiolabelled phage Pf1 and its precursor g5p.DNA complex. Track (a) contains phages isolated from a culture of *P.aruginosum* K grown in 10 ml of MTPA(-Trp) medium with 1 mCi [ $^3\text{H}$ ]-tryptophan. The g3p, g6p, g9p and g7p bands are visible. Track (b) contained g5p.DNA complexes isolated from a similar culture, and shows no corresponding bands. The presence of g3p cannot be ruled out, for the area of the gel near to the origin contained a considerable quantity of radiolabelled protein contaminants. The other minor proteins of the capsid are, however, absent from the g5p.DNA complex.

pooled, diluted tenfold with TE-buffer and sedimented at 100,000 x g for 5 h at 4°C. The high density protein contaminants should have been removed by this stage.

The product was loaded on to a Simons polyacrylamide gel and subjected to electrophoresis and fluorography (Fig. 6.8). A strong band was observed at a migration corresponding with  $M_r$  16,000, approximately that of g5p. Some impurity still remained, but no other bands are visible in the track. This may be compared with the same analysis of Pf1 phages, which shows the presence of g3p, g6p, g7p and g9p. In particular, the absence of g7p and g9p showed that they are not bound in the complex, and thus cannot be pilot proteins directing assembly.

### 6.5 Discussion

The absence of the minor coat proteins of viral capsids from the g5p.DNA complexes of both fd and Pf1 phages is of considerable significance in the understanding of the viral life-cycle, and the regulation of DNA function by proteins in general. Pilot proteins were thought to be polyfunctional phage-coded proteins which guide the DNA through infection, replication and assembly into virions (Kornberg, 1982). Once the four minor protein components of viral capsids were known, the concept of a single pilot protein became unlikely. Instead it might be supposed that each of the minor proteins was responsible for some aspect of the life-cycle. This is in part true, for the g3p of fd has been shown to be responsible for the initial binding of virions to specific hosts (Armstrong *et al.*, 1981).

Since the minor proteins are not present in the g5p.DNA complex, they must be present either in the cytoplasm or the cell membrane awaiting assembly into capsids. The former is unlikely, for they are all rather hydrophobic proteins, which favours the idea of a membrane location, as discussed in Chapter 5. In support of this g3p has been found in the cell membrane and not the cytoplasm of fd-infected cells (Jazwinski et al., 1973), being added to the last end of the particle to be assembled from its prior residence in the membrane (Jazwinski et al., 1973).

The minor proteins are thus very likely to be inserted in the membrane, together with the major capsid protein. Since the g7p and g9p are located at the first-assembled end of the virion, they must be the first to associate with the DNA. They may still have a role in the initiation of assembly, but not in the sense of pilot proteins. At this end of the virion is the A-hairpin of the DNA molecule, which is the first part of the DNA to be assembled into phage. The g7p.g9p is thus likely to be the first component of the capsid to be picked up by the DNA as it passes through the membrane. The orientation of the DNA in the g5p.DNA complex is a problem taken up in the next chapter.



CHAPTER 7

THE DNA OF  
G5P.DNA COMPLEXES

### 7.1 The Possible Orientation of The DNA

The DNA of Ff phages contains regions of potential secondary structure - that is, base-pairing - which are concentrated in the non-coding part of the genome termed the intergenic zone. It is known that the principal (A) hairpin loop exists in the intact virion (the only one to do so) and is located at one end of the particle, where it may be associated with the initiation of assembly (Lopez & Webster, 1984; Armstrong *et al.*, 1981; Ikou & Hearst, 1981). Thus the DNA is orientated within the virion by virtue of the fixed location of this loop. In the previous chapter it was shown that there were no minor coat proteins associated with the g5p.DNA complex of fd or Pf1 viruses, which acts as the immediate precursor of the virus during assembly at the cell membrane. Given that the DNA is orientated within the virion, an obvious question to ask is whether the DNA is orientated specifically within the g5p.DNA complex, and, if so, how. There is at present no information on this point. An obvious guess is that it too is orientated with the strong hairpin loop at one end, ready to enter assembly.

If there is a DNA hairpin at one end of the complex, it is likely that it would not bind to the g5p with the same facility as the remainder of the DNA, which is single-stranded. Therefore the loop might be exposed and be susceptible to digestion with nucleases. The same might be true of the DNA where it turns at the other end of the g5p.DNA complex. Mapping of DNA is normally accomplished with restriction endonucleases, which cut at specific sequences within the DNA. Since the

fd sequence is reliably known (Beck et al., 1978), the pattern of dissection could be predicted. Unfortunately this cleavage is one of double-stranded DNA - and the DNA of g5p.DNA complex is single-stranded apart from the putative hairpin loop.

To overcome this difficulty the following strategy was devised. The nuclease to be used must be capable of cutting single-stranded DNA. It must cut double-stranded DNA only once, and not across both strands, or the frequency of cutting will be artificially high in the loop - though this would not detract from the evidence of that hairpin existing. The enzyme chosen was DNase I, which under suitable conditions fulfils this requirement (Maniatis et al., 1982). After the DNase I digestion, the DNA molecule would be end labelled with  $^{32}\text{P}$  by kinasing and annealed to complete (-) strands, before being subjected to restriction analysis to locate the DNase I cut sites.

## 7.2 Large-scale Preparation of fd g5p.DNA Complexes

In order to supply the quantity of g5p.DNA complex required for the nuclease probing experiment a large-scale isolation of the material was undertaken from phage fd-infected cells. A total culture of 16 l of E.coli K38 was prepared as 10 x 1.6 l of 2xTY medium in 21 flasks. These were inoculated from an overnight culture in 120 ml 2xTY, at 12 ml per flask, and grown with orbital shaking until  $A_{680}$  0.04 was reached, when the culture was infected with phage fd at moi 10. The culture was infected at a lower multiplicity of infection than usual to conserve

phage stocks, and earlier than usual to ensure complete infection nonetheless. After 6 h the cells were harvested by centrifugation at  $10,000 \times g$  for 10 min at  $4^{\circ}\text{C}$ . This took several loads of a Sorvall GS3 rotor, the pellets being accumulated in a beaker kept on ice. A total of 4 g cell paste was recovered and resuspended in 8 l of NT-buffer, centrifuged, resuspended and centrifuged. The final resuspension was in 1 l TE-buffer, which was assayed for  $A_{680}$  and diluted with the same buffer to give  $A_{680}$  of 1.0. Lysis of the complexes was performed with lysozyme (section 2.3.3) and the suspension was cleared by centrifugation at  $15,000 \times g$  for 60 min at  $4^{\circ}\text{C}$ . The large volume of complex bearing fluid was unweildy for ultracentrifugation to sediment the complexes directly, so PEG precipitation was used to concentrate the material (section 6.4). The solution was made 5% in PEG 6000, 0.5 M NaCl, and left in a cold room at  $4^{\circ}\text{C}$  overnight. The precipitate of complexes was collected by centrifugation at  $10,000 \times g$  for 90 min at  $4^{\circ}\text{C}$  in a Sorvall GSA rotor, then resuspended in 10ml TE-buffer and sedimented at  $100,000 \times g$  for 5 h at  $20^{\circ}\text{C}$ . The pellet was resuspended in 0.1 ml TE-buffer and stored at  $4^{\circ}\text{C}$ .

The material contained approximately 10 mg/ml of g5p.DNA complexes, that is, about 1 mg/ml of DNA. Though not pure, it was used as a stock from which small amounts of g5p.DNA complex could be isolated by Metrizamide density gradient centrifugation (section 6.2). Up to 10% of this stock could be run on a 5 ml self-forming gradient of 35% Metrizamide in TE-buffer, to give a pure product of 0.1 mg complex. Samples of this material were run overnight on 1% agarose gels at 1

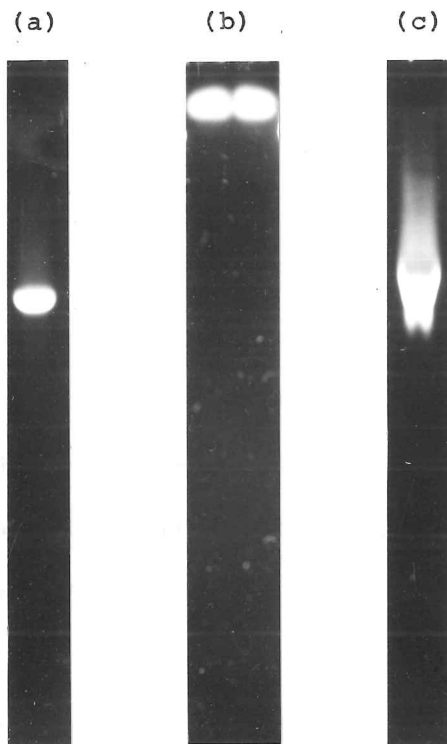


Figure 7.1. The integrity of isolated g5p.DNA complexes shown by agarose gel electrophoresis (section 2.7). Track (a) contained fd DNA isolated from the phage stock by deproteinisation (section 2.6). The complexes in (b) show a restricted electrophoretic mobility, due to their more rigid filamentous structure. Disaggregation of the nucleoprotein in 8 M urea releases a DNA species (c) which comigrates with the fd DNA (though the urea causes the band to "streak" somewhat).

V/cm, showing a restricted mobility relative to pure viral DNA (Fig. 7.1). This indicated that the nucleoprotein filaments were intact. Disaggregation of the preassembly complexes, by incubation with 8 M urea for 30 min at 37°C, freed a DNA species which comigrated with pure phage DNA.

### 7.3 Preparation of Replicative Form DNA

The most readily accessible (-) strands, for the annealing of digested (+) strands before restriction analysis, are those present in the intermediates of viral DNA replication. The (+) strands could be mixed with RF, the whole denatured and then re-annealed. Provided an excess of RF was present, the majority of the digested (+) strands would anneal with (-) stands to give double stranded forms suitable for restriction analysis. The RF forms were prepared as described below, based on a pre-existing protocol (Dr. R.T. Hunt, pers. comm.).

An overnight culture of E.coli K38 was grown in 5 ml 2xTY medium and inoculated into 500 ml of 2xTY, then cultured at 37°C with aerobic orbital shaking until an  $A_{680}$  of 0.34 was reached. The culture was infected with a total of  $5 \times 10^{11}$  phages and grown for 3 h, then 0.67 ml of 30 mg/ml chloramphenicol was added and the culturing continued for 1 hour to allow amplification of the RF. Cells were harvested by centrifugation at  $5,000 \times g$  for 10 min at 4°C, then resuspended in 7.7 ml of 5 mg/ml lysozyme in 50 mM glucose, 25 mM Tris.HCl, 10 mM EDTA, pH 8 and the slurry transferred to a Beckman SW27 tube. After 15 min of

lysis at room temperature, 15.3 ml of 0.2 M sodium hydroxide solution were added, the solutions mixed gently by repeated inversion and allowed to stand in ice for 10 min. Then 11.5 ml of 3 M potassium acetate/acetic acid, pH 4.8, were mixed in vigorously and the tube left for a further 10 min in ice. The precipitate was removed by centrifugation in a Beckman SW 27 rotor at 20,000 rpm for 60 min at 4°C.

The supernatant was decanted into Sorvall GSA rotor bottles, 0.6 volume propan-2-ol added, mixed briefly and allowed to stand at room temperature for 15 min. The DNA precipitate was collected by centrifugation in a GSA rotor at 9,500 rpm for 30 min, the pellet being washed with 70% ethanol in water, then dried briefly under vacuum.

This crude preparation of RF DNA was resuspended in 8 ml TE-buffer, in which were dissolved 8 g caesium chloride and 0.8 ml of 10 mg/ml Ethidium bromide. The solution was centrifuged at 45,000 rpm in a Beckman 50 Ti rotor for 40 h at 20°C. The resulting density gradient separated the DNA into two reddish bands, an upper bacterial DNA band and a lower RF band which was selected. The Ethidium bromide was removed from the RF by extraction against water-saturated N-butanol, repeated four times. The resulting solution was dialysed against 4 batches of TE-buffer for a minimum of 6 h each time.

The suspension of RF thus obtained was precipitated with ethanol. A suitable concentration of monovalent cations was provided by the addition of 1/10th volume of 2.5 M sodium acetate solution.

Precipitation was effected by the addition of 2 volumes ice-cold ethanol to the pre-cooled sample and storage overnight at  $-20^{\circ}\text{C}$ . The DNA was collected by centrifugation at  $10,000 \times g$  for 20 min and resuspended in 1 ml TE-buffer. Samples of this product were subjected to spectrophotometry in an acid-washed quartz cuvette. The stock showed an  $A_{260}$  6.0, indicating about 0.3 mg/ml DNA. The ratio  $A_{260}/A_{280}$  was 1.8, as expected from a pure sample of double-stranded DNA. The product was also shown to be pure by agarose gel electrophoresis: two bands were seen, a heavy band of supercoiled RF I and a faint band of relaxed RF II (Fig 7.2).

Scission with the restriction enzyme Hinc II, which cuts the DNA only once, converted these forms to a band of intermediate mobility, the linear RF III. This is the form needed for the annealing step, for the strands of the supercoiled RF I are topologically linked and would not be able to separate. A sample of 10  $\mu\text{g}$  of the RF preparation was restricted by incubation with 33 units of Hinc II for 3 h at  $37^{\circ}\text{C}$  in medium salt restriction buffer (section 2.5). The solution was deproteinised, and a small sample examined by agarose gel electrophoresis to ensure that linearisation had taken place (Fig. 7.2). This material was to be used in the annealing step.

#### 7.4 Nuclease Digestion of Complexes

With g5p.DNA complex and RF DNA prepared, preliminary digests of free fd DNA were attempted. A DNase I stock was available, at an activity of



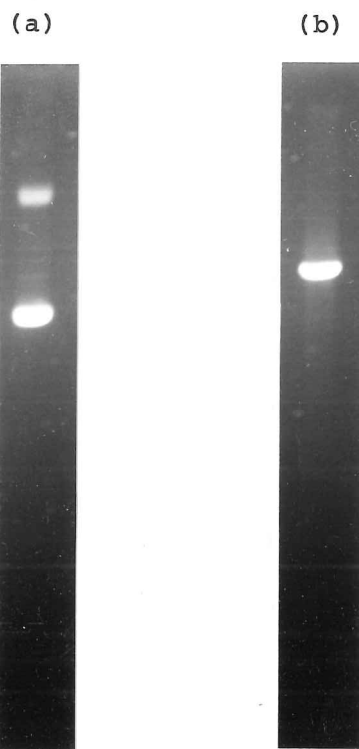


Figure 7.2. Replicative form DNA of fd analysed by agarose gel electrophoresis (section 2.7). Track (a) shows an intense band of RF I (supercoiled duplex DNA) and a fainter band of RF II (relaxed duplex DNA) migrating more slowly. Scission with the restriction endonuclease Hinc II (b) cleaves the DNA at one site only to give the linear duplex form (RF III), of intermediate mobility.

50 units/ $\mu$ l, in 0.1 M hydrochloric acid stored at  $-70^{\circ}\text{C}$ . This was thawed (and subsequently stored at  $-20^{\circ}\text{C}$ ), and diluted tenfold into 10 mM Tris.HCl, pH 7.6, 20 mM  $\text{MgCl}_2$  (DNase I buffer) before use. The diluted sample was allowed to stand for 4 h at  $0^{\circ}\text{C}$  to attain its full activity.

A series of samples of pure viral DNA, obtained by the deproteinisation of phages (section 2.6), were prepared as 1  $\mu$ g DNA in 20  $\mu$ l DNase I buffer. These were incubated with various amounts of DNase I at  $37^{\circ}\text{C}$  for 1 hour, to find an enzyme concentration suitable for gentle lysis of the DNA. The reactions were stopped with 2  $\mu$ l 0.5 M EDTA, and analysed by agarose gel electrophoresis (Fig. 7.3). Gentle degradation was found at a level of 0.1 units enzyme per  $\mu$ g DNA. At higher levels of enzyme the DNA was degraded to oligonucleotides, and at lower levels no cutting was evident.

The only problem remaining was how to make visible the cuts made in the single-stranded DNA once it had been dissected by restriction. Since each cut leaves a 5' and a 3' end, the obvious method of choice was DNA end-labelling. Unfortunately, cutting with DNase I leaves a 5'-phosphate, which must be removed by the action of alkaline phosphatase before radioactive phosphate can be added on from gamma- $^{32}\text{P}$ -ATP by  $\text{T}_4$ -polynucleotide kinase. Subsequently, the DNA would be annealed to the (-) strands of RF and restricted. Restriction fragments should be obtained which are radiolabelled in proportion to the frequency of cutting within their spans of the (+) strand.

(a) (b) (c)

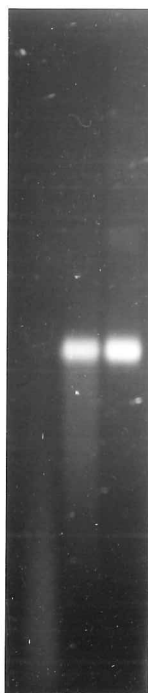


Figure 7.3. DNase I digestion of isolated viral (fd) DNA. Each reaction proceeded for 1 h at 37°C with 1 µg DNA in 20 µl buffer. DNase I was present at (a) 1 unit/µg DNA, (b) 0.1 unit/µg DNA, (c) 0.01 unit/µg DNA. The intermediate concentration was chosen for mild digestion of the g5p.DNA complex.

In order to execute this plan, 2  $\mu$ g of pure complexes (section 7.2) were incubated with 0.2 units of DNase I in 50  $\mu$ l of DNase I buffer for 30 min (equivalent to a digestion of 1 hour with 0.1 units, found above to give gentle degradation). The reaction was stopped with 4  $\mu$ l 0.5 M DNA and the solution was deproteinised by extraction with 'phenol', 'phenol/chloroform' and 'chloroform', and extracted twice against water-saturated ether to remove traces of the denaturants, which might have affected subsequent reactions (section 2.6). Residual ether was blown off under a steady, slow stream of nitrogen for ten minutes.

A sample of cut single-strands of DNA was thus obtained. To this were added 5  $\mu$ l of Sma I buffer (section 2.5) and 20 units of alkaline phosphatase. After a 30 minute incubation at 37°C, another 20 units of the enzyme were added, and the incubation continued for a further 30 min. The resulting solution was extracted as described above to give a deproteinised solution of DNA with free 5'OH ends. To this were added small amounts of concentrated stock solutions to bring the buffer composition to the optimum for the end-labelling reaction: 50 mM Tris.HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 5 mM dithiothreitol, 0.1 mM spermidine, 0.1 mM EDTA with an additional 10 mM KCl from the phosphatase buffer, which should not affect the reaction significantly. The labelling reaction was started by adding 20 units of T<sub>4</sub>-polynucleotide kinase and 10  $\mu$ Ci of gamma-[<sup>32</sup>P]-ATP, and allowed to proceed for 30 min at 37°C. At this point, 3  $\mu$ l 0.5 M EDTA were added to terminate the reaction, and the solution deproteinised as before. The material was recovered in 50  $\mu$ l solution, ready for annealing.

In order to execute this plan, 2  $\mu$ g of pure complexes (section 7.2) were incubated with 0.2 units of DNase I in 50  $\mu$ l of DNase I buffer for 30 min (equivalent to a digestion of 1 hour with 0.1 units, found above to give gentle degradation). The reaction was stopped with 4  $\mu$ l 0.5 M DNA and the solution was deproteinised by extraction with 'phenol', 'phenol/chloroform' and 'chloroform', and extracted twice against water-saturated ether to remove traces of the denaturants, which might have affected subsequent reactions (section 2.6). Residual ether was blown off under a steady, slow stream of nitrogen for ten minutes.

A sample of cut single-strands of DNA was thus obtained. To this were added 5  $\mu$ l of Sma I buffer (section 2.5) and 20 units of alkaline phosphatase. After a 30 minute incubation at 37°C, another 20 units of the enzyme were added, and the incubation continued for a further 30 min. The resulting solution was extracted as described above to give a deproteinised solution of DNA with free 5'OH ends. To this were added small amounts of concentrated stock solutions to bring the buffer composition to the optimum for the end-labelling reaction: 50 mM Tris.HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 5 mM dithiothreitol, 0.1 mM spermidine, 0.1 mM EDTA with an additional 10 mM KCl from the phosphatase buffer, which should not affect the reaction significantly. The labelling reaction was started by adding 20 units of T<sub>4</sub>-polynucleotide kinase and 10  $\mu$ Ci of gamma-[<sup>32</sup>P]-ATP, and allowed to proceed for 30 min at 37°C. At this point, 3  $\mu$ l 0.5 M EDTA were added to terminate the reaction, and the solution deproteinised as before. The material was recovered in 50  $\mu$ l solution, ready for annealing.

One half of the product of the end-labelling reaction was added to 5  $\mu$ g of linearised RF (prepared as described in section 7.3) in a total volume of 100  $\mu$ l. Alkaline denaturation was effected by the addition of 1.5  $\mu$ l 10 M sodium hydroxide, the mixture being allowed to stand for 5 min at room temperature to ensure strand separation. This was followed by neutralisation with 5  $\mu$ l 2.9 M ethanoic acid and 5  $\mu$ l 3 M KCl. Annealing was effected by incubating the solution for 2 h at 65°C and then 30 min at 37°C. The product was chilled on ice, and mixed with 2 volumes of ice-cold ethanol; the DNA was allowed to precipitate overnight at -20°C. Centrifugation at 10,000 x g for 20 min at 4°C gave a pellet of DNA which was resuspended in 40  $\mu$ l of medium salt restriction buffer.

Restriction of the annealed double-strands was performed with 20 units of Hae III for 1 hour at 37°C. The reaction was stopped with 2  $\mu$ l 0.5 M EDTA, and prepared for loading onto an agarose gel by the addition of 4  $\mu$ l of loading buffer (section 2.7). After electrophoresis the restriction pattern was visualised on a UV-transilluminator (Fig. 7.4). The gel was dried down on a water-pump and autoradiographed.

The results showed a highly non-random distribution of radiolabel in the bands (Fig. 7.4). Restriction fragments A and B are the largest and contain little radiolabel, whereas C contains a great deal of label. The fact that different segments of the DNA are labelled to different extents, is proof of their differential exposure to DNase I cutting in the g5p.DNA complex. Evidently, the DNA is not randomly orientated in

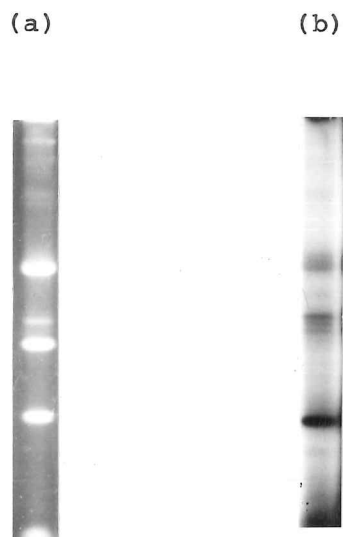


Figure 7.4. The differential susceptibility of the DNA of fd g5p.DNA complexes to digestion with DNase I. After mild digestion with DNase I, the complexes were deproteinised (section 2.6) and the DNA end-labelled with [ $^{32}$ P]-ATP, then annealed with (-) strand DNA and cleaved with restriction endonuclease Hae III. The sample was analysed by agarose gel electrophoresis, and visualised by Ethidium bromide fluorescence (a), and by autoradiography (b). An highly non-random pattern of DNase I scission is shown, for each fragment is labelled in proportion to the frequency of DNase I cutting within its span of the DNA.

the complex, but would appear to have particular sequences exposed at one or both ends where they are susceptible to DNase I. This is in direct parallel with phage capsids, which also contain the DNA in a specific orientation. A reasonable hypothesis would be that these orientations correspond. This was to be examined next, with the same technique extended for much higher resolution.

#### 7.5 The DNase I Susceptibility Pattern of Complex DNA

In order to determine more precisely where the exposed sections of the DNA are on the g5p.DNA complex, the remaining radiolabelled sample was divided into two aliquots of 10  $\mu$ l each and restricted, one with 20 units of Hae III, the other with 20 units of Rsa I, for 1 hour at 37°C. The reactions were stopped with 2  $\mu$ l 0.5 M EDTA and analysed by agarose gel electrophoresis through a 5% gel. This high concentration of agarose was used to ensure resolution of the smaller restriction fragments. Electrophoresis was carried out at 5 V/cm for 5 h, and the band pattern photographed with UV-transillumination, and autoradiographed, as before (Fig 7.5). The results again showed a highly non-random distribution of radioactivity, in both restriction patterns.

The bands thus located were cut out and the agarose removed from the paper backing onto which it had been dried down. The slices taken were 2 mm by 1 cm, and control slices were taken from between the bands. Each slice was placed in 1 ml water in a screw-capped reaction vial, and placed in a boiling waterbath for 10 min to dissolve the agarose and release the DNA. The vials were removed and vortexed thoroughly. Samples of the resulting solution were taken for scintillation counting.



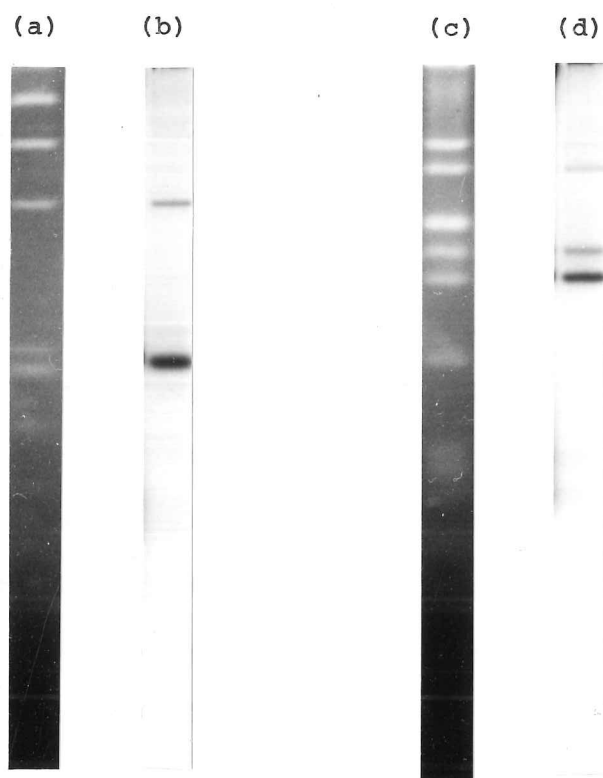


Figure 7.5. The pattern of DNase I susceptibility of the DNA in the g5p.DNA complex of fd, analysed by two separate restriction endonucleases: (a) Hae III, Ethidium bromide stain; (b) Hae III, autoradiography; (c) Rsa I, Ethidium bromide stain; (d) Rsa I, autoradiography. A highly non-random distribution of radiolabel is visible in both cases. The DNA must be present in a specific linear orientation within g5p.DNA complexes, certain sequences being exposed at the ends of the filament and thus more susceptible to DNase I scission.

The distribution of radiolabelling, hence susceptibility to DNase I, was plotted as a histogram along the genome (Fig. 7.6). As the histograms show, there are two zones of exposed DNA in the g5p.DNA complex, one around the area of the intergenic zone, and another in gene 3, opposite the intergenic zone on the circular genetic map. If these two patterns are taken as estimates of the true probability distribution for DNase I cutting, they may be combined to give a better estimate of that distribution. Essentially the DNA is divided up into class intervals at all the restriction sites considered, as if it had been digested with both restriction enzymes. For each fragment the best estimate of the probability of a cut within that span, is the mean of the probabilities estimated by each of the observed restriction patterns (J.M.C. Rawlings, pers. comm.).

The result of this procedure is shown in Figure 7.7. The two exposed zones of DNA in g5p.DNA complexes are clearly visible as peaks in the probability distribution for nuclease cutting. Their distance apart and the considerations of symmetry in the g5p.DNA complex mean that they must be at opposite ends of the complex. The summit of the taller peak covers the range 5519 to 5831 nucleotides on the DNA sequence filed at this laboratory. This covers most of the intergenic zone (5503 to 6126). A control experiment was performed, in which the same experimental procedure was used to determine the DNase I susceptibility pattern of free fd DNA. A random distribution of cutting was observed (Fig. 7.8).

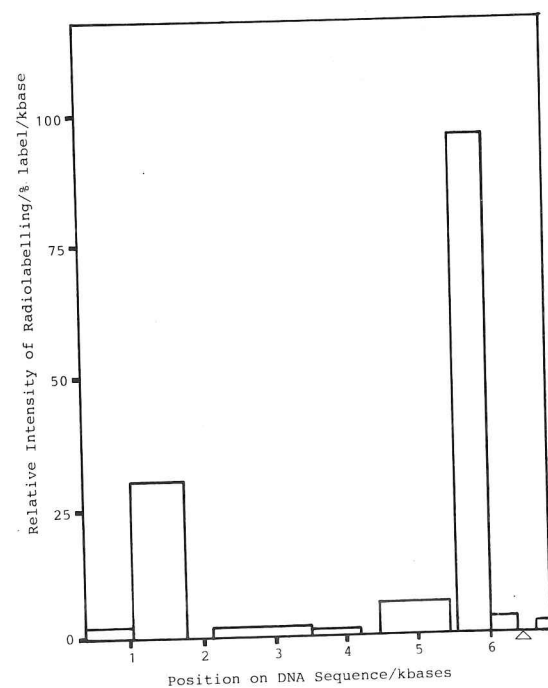
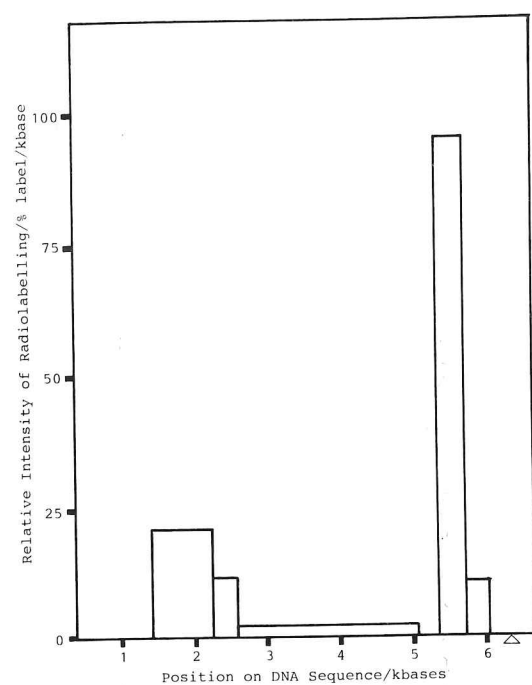


Figure 7.6. The distribution of radioactivity, hence of DNase I cutting in the g5p.DNA complex, plotted along the length of the fd genome. Above: Hae III analysis. Below: Rsa I analysis. Two peaks appear to be present, one around the intergenic zone and another some distance away from it. The zones of preferential susceptibility presumably lie at the two ends of the filamentous g5p.DNA complex

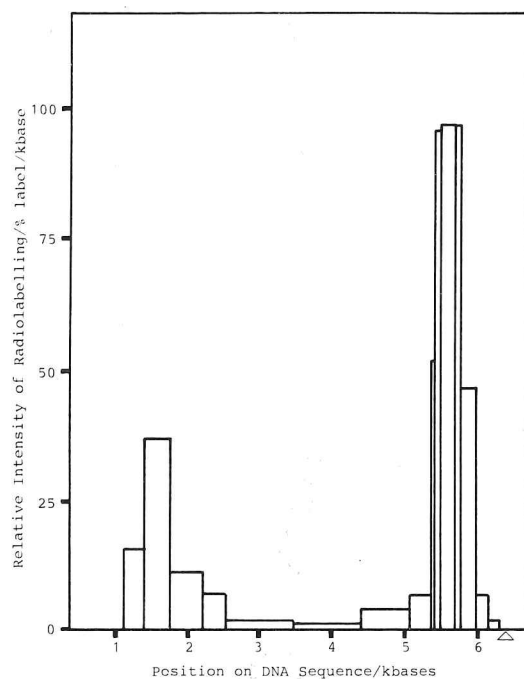


Figure 7.7. The information of the two histograms in Figure 7.6 may be combined to give a clearer picture of the pattern of DNase I susceptibility in the DNA of g5p.DNA complexes. The taller peak covers the range 5519 to 5831 nucleotides, corresponding approximately with the intergenic zone (5503 to 6126 nucleotides).

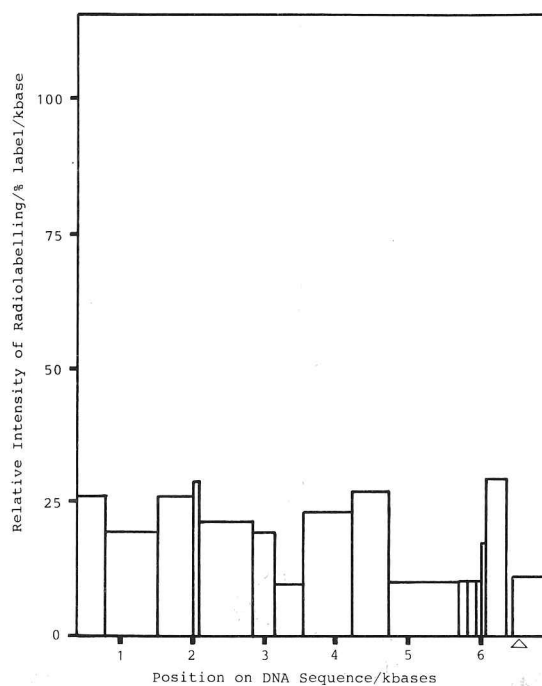


Figure 7.8. The pattern of DNase I susceptibility of free fd DNA, plotted along the length of the fd genome. An effectively random distribution is observed, indicating that the modulation of the susceptibility visible in Figure 7.7 is due to the protein of the g5p.DNA complex.

The whole experiment was also performed using three separate restriction enzymes to dissect the DNA from DNase I digestion of g5p.DNA complexes. A sample of 1 µg pure complex was digested with 0.02 units of DNase I for 30 min at 37°C. The subsequent analysis of the digestion was as performed previously, except that three restriction enzymes were used to separately restrict the annealed DNA: Hae III, Rsa I and Taq I. The DNase I susceptibility pattern of fd DNA was thus obtained at even higher resolution (Fig. 7.9).

The two peaks in the distribution are clearly visible. The sharper one has its maximum at 5465 to 5686 nucleotides, comprising 3% of the genome, and the broader one at 1510 to 1951 residues, comprising 7% of the genome. The sharp peak falls within the intergenic zone, the most exposed region being centred on the A-hairpin (Fig. 1.5). This is consistent with the existence of this hairpin as a protein-free, apical structure in g5p.DNA complexes. The A-hairpin is the first part of the DNA to be encapsidated during assembly (Lopez & Webster, 1982), so the initial interaction of the morphogenetic process must be between this exposed region of DNA and capsid proteins, probably bound in the bacterial membrane. The initiator of phage morphogenesis is not, therefore, a protein signal but the intergenic DNA exposed from one end of the g5p.DNA complex. The peak of cutting probability roughly antipodal to the intergenic zone is more diffuse suggesting that the exposed frame may be less precise than at the A-hairpin end. This may be a consequence of the structure of the g5p.DNA complex at this end.

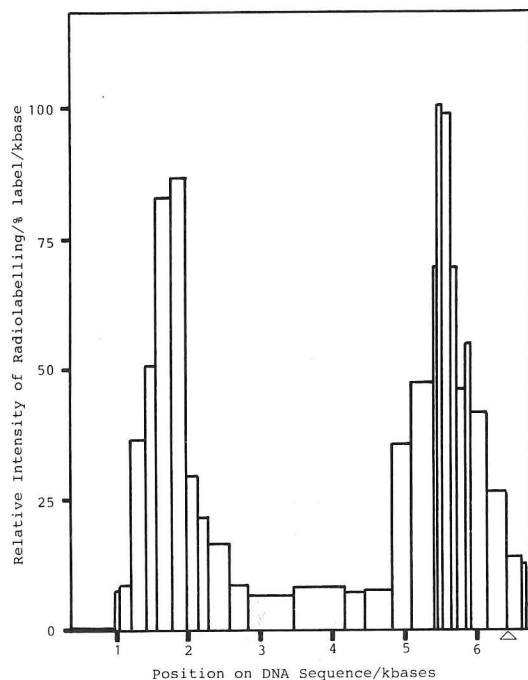


Figure 7.9. The distribution of DNase I susceptibility of the DNA in the g5p.DNA complex plotted along the length of the fd genome. To get such a finely spaced distribution the results from three restriction endonuclease mappings (Hae III, Rsa I and Taq I) were combined. The sharper peak has a maximum at 5465 to 5686 nucleotides, inside the intergenic zone and corresponding with a small region of DNA including the whole of the A-hairpin (5504 to 5579 nucleotides). The broader peak, from the susceptibility of the DNA at the other end of the complex, has its maximum between 1510 and 1951 nucleotides.

In fact, the DNA exposed at this end of the complex is not precisely opposite the intergenic zone on the genomic map (Fig. 7.10): that would require the peak to be centred on 2336 residues, whereas it is actually centred on 1730. This is a difference of 606 nucleotides, about 9% of the genome. Why should this be? One possibility is that the DNA follows a longer path one way down the complex than the other, and that it is always the same strand which follows the shorter length path. In a helical complex this situation can be produced readily if, overall, one strand of the DNA forms a helix at a greater radius from the central axis than the other. This gives the outer DNA strand a longer path to traverse in the same number of helical turns than the strand following the tighter helix. The concept is illustrated in Figure 7.11.

An interesting speculation based on this model is that the shorter path DNA must be assembled into the g5p.DNA complex first. This requirement is due to the fact that once g5p has bound to the DNA, local protein-protein interactions force it into a helical array, concealing the inner path binding sites. Thus the inner helix might be filled with DNA first, then the outer helix wound around the assembled complex. The reverse is unlikely, as it would involve DNA being fed up the axial hole in the complex to fill the empty binding sites.

In fact, the shorter path DNA from the A-hairpin to the antipodally exposed region runs in the direction 5' to 3', that is, in the direction of DNA synthesis. Thus the DNA to be first synthesised is that found in the shorter path helix. The correct assembly of the g5p.DNA complex has



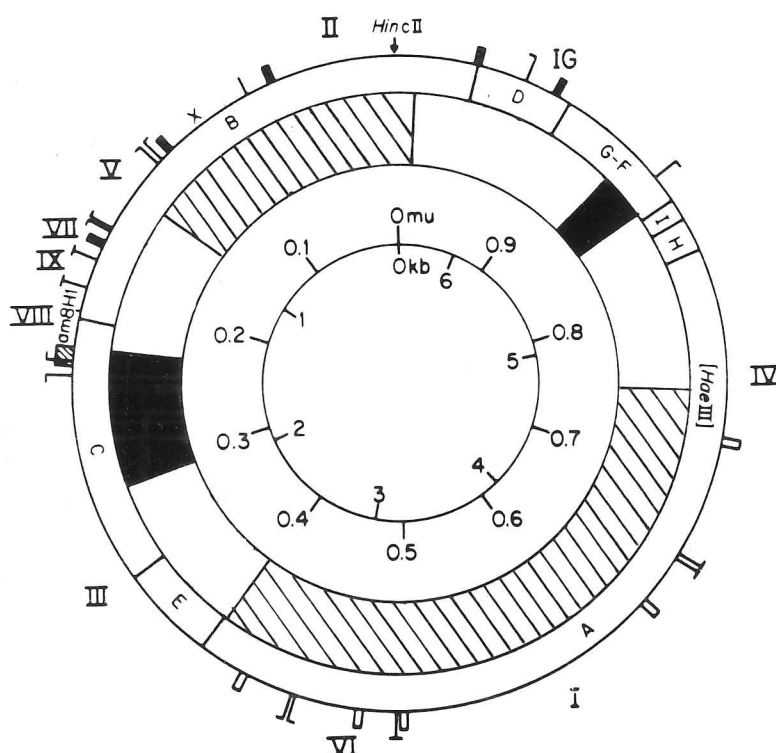


Figure 7.10. The pattern of DNase I susceptibility of the DNA of g5p.DNA complexes shown on the genetic map (after Moses *et al.*, 1980). Regions of particularly high susceptibility (greater than 30% total label/kbase) are shown in black. Regions of particularly low susceptibility (less than 5% total label/kbase) are shaded, fitting neatly between the regions of high susceptibility. The latter are not directly opposite on the map, and one possible reason for this is discussed in the text.

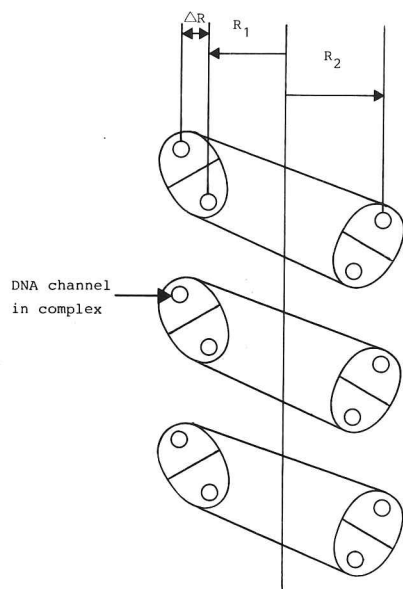


Figure 7.11. A possible structure for the helical g5p.DNA complex, viewed as a cutaway from the side. If the DNA follows two paths of differing lengths between the ends of the complex, this may be because their helical paths are of different radii. A 9% path difference may be accounted for by a 18% difference between  $R_1$  and  $R_2$  as illustrated.

been known to depend on DNA synthesis for some time (Zentgraf et al., 1977). This is easily explained if the complex is formed initially between g5p and the first synthesised DNA of the phage genome. In the course of (+) strand synthesis this is the part of the DNA beginning at the intergenic zone and extending 3'-wards (Horiuchi et al., 1978; Horiuchi & Zinder, 1976). Once g5p dimers have bound to the DNA they undergo lateral interactions to form a loose helix, the later synthesised DNA being wound around the outside.

In summary, then, the g5p.DNA complex of fd was shown to have two regions of DNA preferentially susceptible to DNase I digestion, presumably at either end of the particle. One, a sharply defined region, corresponds with the A-hairpin of the intergenic zone and a little DNA on either side of it. This is envisaged as the signal initiating morphogenesis of phages, interacting with the capsid proteins waiting in the membrane. The helical complex appears to contain distinct "up" and "down" stands of DNA, with different lengths, the structure being consistent with the assembly of the inner helix first during DNA synthesis.

## CHAPTER 8

### DISCUSSION

### 8.1 The Structures of Filamentous Phages

At this stage, the current working models of filamentous phage architecture will be discussed. In particular, the modifications which have been made to the models in the course of this work will be considered in the perspective of the literature on the subject.

The protein components of filamentous bacteriophages from both Classes I and II show considerable and unexpected parallels, despite the very different arrangements of their major capsid proteins. This difference in symmetry, together with the differences in nucleotide to subunit stoichiometry (Newman et al., 1977; Berkowitz & Day, 1976), led to the feeling that Class I and Class II phages were fundamentally different in construction (Makowski et al., 1980; Wiseman & Day, 1977; Wiseman et al., 1976; Marvin et al. 1974a; Marvin et al., 1974b). The finding that examples of each class contain corresponding sets of major and minor protein components highlights considerable homology of capsid structure (Chapters 3 and 4). The case is strengthened by the fact that the minor protein components of both classes are present in the same relative stoichiometries, within the limits of experimental error - 1:1:1:1 for g3p/g6p/g7p/g9p. This calls into question the idea that the capsids have a fundamentally different architecture. The differences in symmetry reflect different modes of assembling a protein tube. Surely this is out-weighed by the close correspondence in capsid components, and their presumed functions in capsid assembly and host infection.

Analysis of the potential secondary structures of the capsid proteins showed some parallels between those of the Class I phage fd and the Class II phage Pf1 (Chapter 5). These analyses were based on the amino-acid sequences translated from the putative coding regions of the fd and Pf1 DNA (Beck et al., 1978; N.J. Short, pers. comm.), and the recognition of these proteins in the Pf1 capsid (Chapter 4) strengthens the identification of their respective genes in the Pf1 DNA sequence. The predicted secondary structures of g8p, g7p and the parts of g3p analysed show substantial homology, whereas g6p and g9p are less clearly related. A very close evolutionary relationship between the two classes of phage is indicated.

On the other hand, divergence of the g8p of fd and Pf1 from a common ancestor has been placed at  $3 \times 10^9$  years before present, on the basis of differences in the known amino-acid sequences (Nakashima et al., 1975; Nakashima et al., 1974). This figure is suspect, because the phage reproductive cycle is about 20 minutes, very much shorter than that of the "typical organism" it was assumed to be. Also, being very simple systems, relative to cellular organisms, filamentous phages can undergo adaptation more rapidly: changing a few amino-acids in a capsid protein constitutes a large proportionate change of the genome. It is not unlikely that divergence was much more recent, and followed by rapid adaptive evolution. Indeed, some evolution is observed in serial passages of viruses through several culture cycles (Folkhard et al., 1979). The most common products are strains of phage which contain an incomplete genome (miniphage, or defective interfering particles - DIP's), but manage to replicate by co-infecting cells with wild-type phages (Enea & Zinder, 1982; Griffith & Kornberg, 1974).

## 8.2 The Nucleoproteins of the Life-cycle

The life-cycles of the filamentous phages are characterised by a series of nucleoprotein complexes which direct the DNA through the various stages. Although sought, no evidence was found of pilot proteins (Kornberg, 1974) directing this process. The virion itself is a complex directing the DNA towards the infection of a host cell. This function is mediated in the first instance by g3p, which adsorbs onto the receptors (pili) of the host cell membrane (Armstrong et al., 1981). Subsequently the disassembly of the g8p subunits is responsible for releasing the DNA to the cell cytoplasm. This is a linear disassembly process, and can be split into initiation, propagation and termination phases. Initiation must be mediated by the proteins at the g3p.g6p-end of the virion, as this is the first part to disassemble. Since the g3p is known to have a good membrane retention sequence, and is found in the membranes of infected cells (Boeke & Model, 1982) the approximation of the bacterial membrane may allow it to flip from the capsid into that membrane.

The propagation of disassembly proceeds as g8p relaxes from its alpha-helical capsid state to the membrane bound form of mixed conformation (alpha-helix and beta-pleat) (Nozaki et al., 1978). Similarly, membrane insertion sequences of the minor proteins should allow them to "dissolve" in the membrane (sections 5.3-5.6). Termination of disassembly probably involves no specific signal - the process runs up to the end of the capsid.

The next nucleoprotein complex involved in the life-cycle is that between the released DNA and the host SSB protein. This pre-initiation complex directs the DNA towards replication. The complex is slightly unusual as it involves host cell proteins, and is circular not linear. The region of secondary structure at the intergenic zone is left exposed for replicative enzymes to initiate DNA synthesis at that location (Gray *et al.*, 1978; Schaller, 1978; Geider & Kornberg, 1974; Sigal *et al.*, 1972).

In eventual competition with the SSB for free (+) strand DNA is the phage encoded g5p. This binds the single-strand to give a helical nucleoprotein complex directing the DNA towards assembly, the g5p.DNA complex (Pretorius *et al.*, 1975, Pratt *et al.*, 1974; Pratt & Erdahl, 1968). A suggestion of long standing in this field was that proteins present in a few copies - pilot proteins - might guide the DNA through the replicative cycle and assembly into virions (Kornberg, 1974). The g5p.DNA complexes of both the Class I phage fd and the Class II phage Pf1 were shown not to contain any minor protein components which could perform such functions (Chapter 6). In particular, neither g3p, initially suggested as a pilot protein, nor g7p.g9p, which could have been alternatives in the light of more recent research, were detected. The pilot protein concept is therefore not applicable to these systems.

The g5p acts as a pre-forming scaffold for the DNA, winding the single-strand before it is assembled into virions (Kneale & Marvin, 1984). Part of the intergenic zone, containing the A-hairpin, was shown



to be exposed at one end of the g5p.DNA complex (Chapter 7). This zone is also terminal in the virion (Ikoku & Hearst, 1981; Webster et al., 1981; Shen et al., 1979), and exposed in the pre-initiation complex (Geider & Kornberg, 1974), and seems to be a primary organising centre for the important nucleoprotein complexes of the life-cycle. Since it is apical both in phage and complex, and the first part of the virion to be assembled (Lopez & Webster, 1984; Armstrong et al., 1981), this region must contain the signal which initiates phage morphogenesis. An initial interaction between the A-hairpin and the g7p (and perhaps g9p) could be envisaged as the trigger for phage assembly at the membrane (Fig 8.1).

As a linear structure, the g5p.DNA complex itself must be assembled. Only DNA coming straight from the replicative cycle gives rise to the normal linear complexes observed in vivo, with the correct subunit to nucleotide ratio. Reassociation of the components in vitro gives complexes of altered morphology and stoichiometry (Zentgraf et al., 1977; Anderson et al., 1975; Pretorius et al., 1975). The initiation of assembly is therefore not a simple recognition of a specific sequence: DNA synthesis starts in the intergenic zone on hairpin D (Fig. 1.5) and the complex evidently assembles with this region at one end. Consequently the DNA of complexes is in a specific linear register, as was shown, and runs the length of the complex in two concentric helices. This is in accord with the high-resolution structure of the complex derived by Gray et al. (1982), which shows an axially flattened thread. It is interesting to note that the A-hairpin is in the last 300 nucleotides to be polymerised. Thus, any interaction between this region of the DNA and capsid proteins cannot occur until the g5p.DNA complex is almost complete.

## Bacteriophage fd Assembly

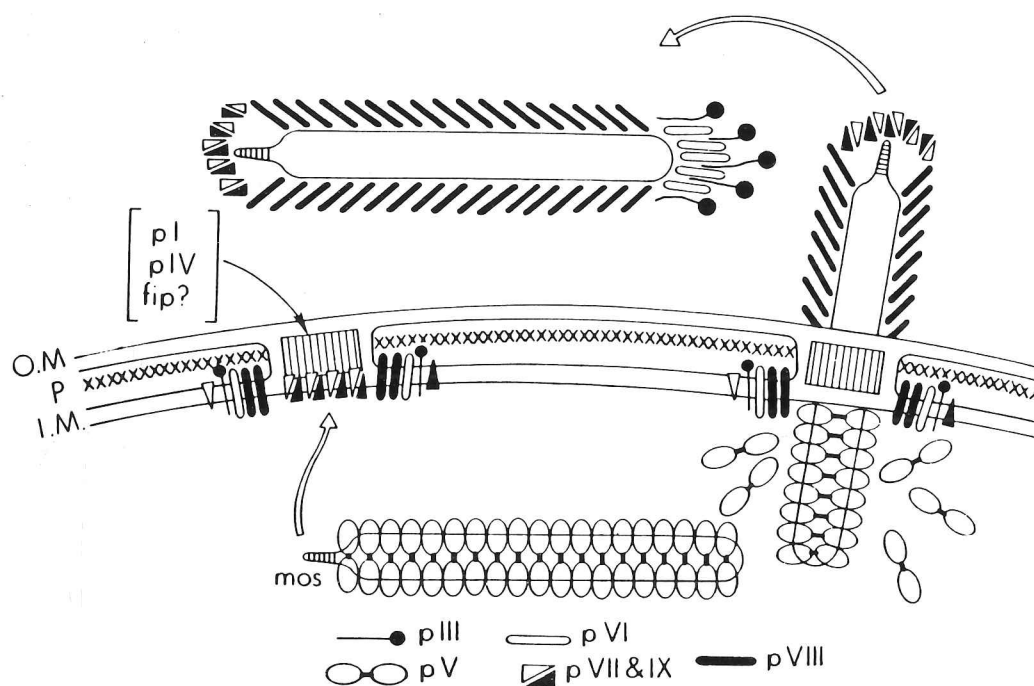


Figure 8.1. The assembly of filamentous bacteriophages from their precursor complexes. Both fd and Pf1 capsids were shown to contain minor protein components (g3p, g6p, g7p and g9p) which were shown to be absent from the precursor g5p.DNA complexes. The properties of these minor proteins indicate that they may be awaiting assembly in the membrane. The DNA of the g5p.DNA complex has been shown to be linearly oriented in the same way as in the phage capsid, with a small region of DNA around the A-hairpin exposed, presumably at one end of the filament. Phage Pf1 is likely to follow the same pattern.

(After Lopez & Webster, 1984)

Subsequently, the DNA picks up g8p from the membrane while shedding g5p to the cytoplasm in a concerted disassembly/assembly process. If there are no specific DNA-g8p interactions, there is no obvious reason for the DNA to be extruded with the coat protein (Berkowitz & Day, 1976). The initial association of the A-hairpin with the g7p.g9p could explain how the DNA is pulled from the cytoplasm, if it is anchored at the distal end of the elongating capsid.

### 8.3 Future Perspectives

In the whole life cycle one thing is clear: the crucial role of the intergenic zone in controlling various stages of the cycle. It remains free in the pre-initiation complex, in which it directs the priming reaction which commits the DNA to replication (Dotto et al., 1982; Geider et al., 1978); free again in replication it contains both the (+) and (-) stand origins; free again (as shown in this work) in the g5p.DNA complex, it is the first-assembled part of the virion, and hence probably the initiator of assembly.

It would be of interest to determine the structural arrangement of this zone in the virion. As it is known that the intergenic zone is always at the g7p.g9p-end of the capsid, it is possible that some specific interactions occur between the secondary structures of this zone and the adjacent proteins. Crosslinking of the DNA with these proteins, using UV-light and possibly the crosslinking agent p-azidophenylglyoxal, may reveal the specific abutments of the proteins on the DNA. Not only might the relevant protein be identified, but also its binding zone on the DNA.

The techniques of molecular biology could be applied usefully to this work. A Pf1 phage constructed with with the fd g8p could have fd symmetry (if the protein is the determining factor), Pf1 symmetry (if the DNA or minor proteins control the symmetry), or some completely new symmetry. Genetic interchange of the g3ps may give rise to fd phages infectious to the Pf1 host, and vice versa. This would be the first example of an engineered change in the host specificity of a virus. The techniques of site-directed mutagenesis would allow more refined manipulations to be carried out. Specific changes in the amino-acid sequences of the proteins should show up the active residues. Such work is in progress (Dr. G. Hunter & D. Rowitch, pers. comm.) to analyse potential interactions between amino-acids near the C-terminal end of fd g8p (the end abutting on the DNA) and encapsidated DNA. Similar probing of the minor coat proteins could throw light on the functionally active residues.

The g5p.DNA complex too, presents opportunities for the study of nucleoprotein complexes. The observation that the DNase I susceptible zones are not exactly opposed on the genetic map indicates that the DNA follows two slightly separated paths in the complex. X-ray diffraction of oriented complexes, or crystallised complex fragments, should reveal how this comes about.

It is generally axiomatic that investigations should start with simple systems and work towards more complex ones. The filamentous bacteriophages are simple viruses. They are hazard-free, readily

prepared systems to which a great variety of physico-chemical techniques both can be and have been applied. In their life-cycles they exhibit a diverse pattern of nucleoprotein complexes which control the fate of the DNA. Both from the point of view of the virologist and of the structural biochemist, this system provides possibilities for productive study far from exhausted.

## References

- Alma, N.C.M., Harmsen, B.J.M., de Jong, E.A.M., v.d. Ven, J. & Hilbers, C.W. (1983) *J. Molec. Biol.* 163 47-62
- Amako, K. & Yasunaka, K. (1977) *Nature* 267 862-863
- Anderson, T.F. (1958) *Cold Spring Harbor Symp. Quant. Biol.* 23 47-58
- Anderson, R.A., Nakashima, Y. & Coleman, J.E. (1975) *Biochemistry* 14 907-917
- Arber, W. (1966) *J. Molec. Biol.* 20 483-496
- Armstrong, J., Perham, R.N. & Walker, J.E. (1981) *Febs. Letts.* 135 167-172
- Asbeck, F., Bayreuther, K., Kohler, H., von Wettstein, G. & Braunitzer, G. (1969) *Hoppe-Seyler's Z. Physiol. Chem.* 350 1047-1066
- Atha, D. & Ingham, K. (1981) *J. Biol. Chem.* 256 12108-12117
- Bailey, G.S., Gillet, D., Hill, D.F. & Petersen, G.B. (1977) *J. Biol. Chem.* 252 2218-2225
- Bailey, G.S., Hill, D.F., Gillet, D. & Petersen, G.B. (1975) *Proc. Univ. Otago Medical School* 53 3-4
- Banner, D.W., Nave, C. & Marvin, D.A. (1981) *Nature* 289 814-816
- Beck, E., Sommer, R., Auerswald, E.A., Kurz, C., Zink, B., Osterbug, G. & Schaller, H. (1978) *Nucleic Acids Res.* 5 4495-4503
- Bendet, I.J. & Mayfield, J.E. (1967) *Biophys. J.* 7 111-119
- Berkowitz, S.A. & Day, L.A. (1976) *J. Molec. Biol.* 102 531-547
- Blobel, G. & Doberstein, B. (1975) *J. Cell Biol.* 67 835-851
- Boeke, J.D. & Model, P. (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79 5200-5204
- Bradley, D.E. (1964) *J. Gen. Microbiol.* 35 471-487
- Bradley, D.E. & Dewar, C.A. (1967) *J. Gen. Virol.* 1 179-188
- Brinton, C.C., Gemski, P. & Charnan, J. (1964) *Proc. Natl. Acad. Sci. U.S.A.* 52 776-783

- Brown, L.R. & Dowell, C.E. (1969) J. Virol. 2 1290-1295
- Brutlag, D., Scheckman, R. & Kornberg, A. (1971) Proc. Natl. Acad. Sci. U.S.A. 68 2826-2829
- Butler, P.J.G. (1984) J. Gen. Virol. 65 253-279
- Caro, L.G. & Schnos, M. (1966) Proc. Natl. Acad. Sci. U.S.A. 56 126-132
- Caspar, D.L.D. & Klug, A. (1962) Cold Spring Harbor Symp. Quant. Biol. 27 1-24
- Caspar, D.L.D. & Makowski, L. (1981) J. Molec. Biol. 145 611-617
- Cavalieri, S.J., Neet, K.E. & Goldthwaite, D.A. (1976) J. Molec. Biol. 102 697-711
- Chamberlain, B.K., Nozaki, Y., Tanford, C. & Webster, R.E. (1978) Biochim. Biophys. Acta 510 18-37
- Chamberlain, B.K. & Webster, R.E. (1976) J. Biol. Chem. 251 7739-7745
- Chan, T-S., Model, P. & Zinder, N.D. (1975) J. Molec. Biol. 99 369-382
- Chang, C.N., Model, P. & Blobel, G. (1979) Proc. Natl. Acad. Sci. U.S.A. 76 1251-1255
- Chang, C., Weiskopf, M. & Li, H.J. (1973) Biochemistry 12 3028-3032
- Chou, D.Y. & Fasman, G.D. (1974) Biochemistry 13 211-222
- Chun, P.W., Herschleb, W.P., Buckley, R.L., Blitziotes, M.M., Son, U.T. & Krista, M.L. (1974) Biochem. Biophys. Res. Comm. 56 907-913
- Cleary, J.M. & Ray, D.S. (1980) Proc. Natl. Acad. Sci. U.S.A. 77 4638-4642
- Cleary, J.M. & Ray, D.S. (1981) J. Virol. 40 197-203
- Cross, T.A. & Opella, S.J. (1982) J. Molec. Biol. 159 543-549
- Cross, T.A. & Opella, S.J. (1985) J. Molec. Biol. 182 367-381
- Date, T. Goodman, J.M. & Wickner, W.T. (1980a) Proc. Natl. Acad. Sci. U.S.A. 77 4669-4673
- Date, T. & Wickner, W. (1981) Proc. Natl. Acad. Sci. U.S.A. 78 6106-6110

- Date, T., Zwizinski, C., Ludmerer, S. & Wickner, W. (1980b) Proc. Natl. Acad. Sci. U.S.A. 77 827-831
- Day, L.A. (1966) J. Molec. Biol. 15 395-398
- Day, L.A. (1969) J. Molec. Biol. 39 265-277
- Day, L.A. (1973) Biochemistry 12 5329-5339
- Day, L.A., Wiseman, R.L. & Marzec, C.J. (1979) Nucleic Acids Res. 7 1393-1403
- Dimicoli, J.L. & Helene, C. (1974) Biochemistry 13 714-723
- Dotto, G.P., Enea, V. & Zinder, N.D. (1981a) Proc. Natl. Acad. Sci. U.S.A. 78 5421-5424
- Dotto, G.P., Enea, V. & Zinder, N.D. (1981b) Virology 114 463-473
- Dotto, G.P. & Horiuchi, K. (1981) J. Molec. Biol. 153 169-176
- Dotto, G.P., Horiuchi, K., Jakes, K.S. & Zinder, N.D. (1982) J. Molec. Biol. 162 335-343
- Dotto, G.P., Horiuchi, K. & Zinder, N.D. (1984) J. Molec. Biol. 172 507-522
- Dunker, A.K., Williams, R.W. & Peticolas, W.L. (1979) J. Biol. Chem. 254 6444-6448
- Edens, L., Konings, R.N.H. & Schoenmakers, J.G.G. (1978) Virology 86 354-367
- Enea, V. & Zinder, N.D. (1982) Virology 122 222-226
- Falaschi, A. & Kornberg, A. (1965) Proc. Natl. Acad. Sci. U.S.A. 54 1713-1720
- Fareed, G., Ippen, K.A. & Valentine, R.A. (1966) Biochem. Biophys. Res. Comm. 25 275-284
- Fersht, A. (1985) Enzyme Structure and Mechanism (pub. W.H. Freeman)
- Fidanián, M.M. & Ray, D.S. (1972) J. Molec. Biol. 72 51-63
- Folkhard, W., Leonard, K.R., Malsey, L.S., Marvin, D.A., Dubochet, J., Engel, A., Achtman, M. & Helmuth, R. (1979) J. Molec. Biol. 130 145-160



Fraenkel-Conrat, H. & Williams, R.C. (1955) Proc. Natl. Acad. Sci.

U.S.A. 41 690-698

Frank, H. & Day, L.A. (1970) Virology 42 144-154

Gabbay, E.J., Sandford, K., Baxter, C.S. & Kpicak, L. (1973)

Biochemistry 12 4021-4029

Geider, K., Beck, E. & Schaller, H. (1978) Proc. Natl. Acad. Sci.

U.S.A. 75 645-649

Geider, K. & Kornberg, A. (1974) J. Biol. Chem. 249 3999-4005

Geider, K. & Meyer, T.F. (1978) Cold Spring Harbor Symp. Quant. Biol.

43 59-62

Goldsmith M.E. & Konigsberg, W.H. (1977) Biochemistry 16 2686-2694

Goodman, J.M., Watts, C. & Wickner, W. (1981) Cell 24 437-441

Grant, R.A., Lin, T-C., Konigsberg, W. & Webster, R.E. (1981) J. Biol.

Chem. 256 539-546

Gray, C.W., Brown, R.S. & Marvin, D.A. (1981) J. Molec. Biol. 146

621-627

Gray, C.W., Kneale, G.G., Leonard, K.R., Siegrist, H. & Marvin, D.A.

(1982a) Virology 116 40-52

Gray, D.M., Gray, C.W. & Carlson, R.D. (1982b) Biochemistry 21

2702-2713

Gray, C.P., Sommer, R., Polke, C., Beck, E. & Schaller, H. (1978) Proc.

Natl. Acad. Sci. U.S.A. 75 50-53

Griffith, J. & Kornberg, A. (1974) Virology 59 139-152

Griffith, J., Manning, M. & Dunn, K. (1981) Cell 23 747-753

Harrison, S.C. (1983) J. Molec. Biol. 171 577-580

Harth, G., Baümel, I., Meyer, T.F. & Geider, K. (1981) Eur. J.

Biochem. 119 663-668

Helene, C., Montenay Garestier, T. & Dimcoli, J.L. (1971) Biochim.

Biophys. Acta 254 349-365

- Henry, T.J. & Pratt, D. (1969) Proc. Natl. Acad. Sci. U.S.A. 62  
800-807
- Heurman, R., Neugebauer, K., Zentgraf, H. & Schaller., H. (1978) Molec.  
Gen. Genetics 159 171-178
- Hill, D.F. & Petersen, G.B. (1978) Proc. Univ. Otago Medical School  
56 52-54
- Hill, D.F. & Petersen, G.B. (1982) J. Virol. 44 32-46
- Hinz, H.-J., Greulich, K.O., Ludwig, H. & Marvin, D.A. (1980) J. Molec.  
Biol. 144 281-289
- Hoffman-Berling, H., Kaerner, H.C. & Knippers, R. (1966) Adv. Virus  
Res. 12 329-370
- Hoffman-Berling, H. & Mazé, R. (1964) Virology 22 305-313
- Hofschneider, P.H. & Preuss, A. (1963) J. Molec. Biol. 7 450-451
- Holmes, K.C. (1982) Structural Molecular Biology, Ed. Davies, D.B.,  
Saenger, W. & Danylu, K.S.S. (Plenum press) 475-505
- Horne, R.W. & Wildy, P. (1961) Virology 15 348-373
- Horiuchi, K. (1980) Proc. Natl. Acad. Sci. U.S.A. 77 5226-5229
- Horiuchi, K. (1984) J. Molec. Biol. 169 389-407
- Horiuchi, K., Ravetch, J.V. & Zinder, N.D. (1978) Cold Spring Harbor  
Symp. Quant. Biol. 389-399
- Horiuchi, K. & Zinder, N.D. (1975) Proc. Natl. Acad. Sci. U.S.A. 72  
2555-2558
- Horiuchi, K. & Zinder, N.D. (1976) Proc. Natl. Acad. Sci. U.S.A. 73  
2341-2345
- Huang, G.C. & Hearst, J.E. (1981) Nucleic Acids Res. 9 5587-5599
- Hulsebos, T. & Schoenmakers, J.G.G. (1978) Nucleic Acids Res. 5  
4677-4698
- Ikehara, K. & Utiyama H. (1975) Virology 66 316-321
- Ikoku, A.S. & Hearst, J.E. (1981) J. Molec. Biol. 151 245-259
- Inman, R.B. (1967) J. Molec. Biol. 25 209-216

- Irie, S. & Sezaki, M. (1983) Anal. Biochem. 134 471-478
- Ito, K., Date, T. & Wickner, W. (1980) J. Biol. Chem. 255 2123-2130
- Jacobson, A. (1972) J. Virol. 10 835-842
- Jazwinski, S.M., Marco, R. & Kornberg, A. (1973) Proc. Natl. Acad. Sci. U.S.A. 70 205-209
- Kaguni, J. & Ray, D.S. (1979) J. Molec. Biol. 135 863-878
- Katsura, I. (1984) J. Molec. Biol. 171 297-317
- Klug, A. (1983) Nature 303 378-379
- Kneale, G.G. (1983) Biochim. Biophys. Acta 739 216-224
- Kneale, G.G. & Marvin, D.A. (1982) Virology 116 53-60
- Kneale, G.G. & Marvin, D.A. (1984) J. Molec. Biol. 171 229-232
- Kneale, G.G. & Tsugita, A. (1985) Biochem. J. 228 193-199
- Knight, C.A. (1974) Molecular Virology (pub. McGraw-Hill)
- Knippers, R. & Hoffman-Berling, H. (1966a) J. Molec. Biol. 21 281-292
- Knippers, R. & Hoffman-Berling, H. (1966b) J. Molec. Biol. 21 293-304
- Knippers, R. & Hoffman-Berling, H. (1966c) J. Molec. Biol. 21 305-312
- Knoll, D. & Herman, J. (1983) J. Biol. Chem. 258 5710-5715
- Kochan, J., Carrascosa, J.L. & Murialdo, H. (1984) J. Molec. Biol. 174 433-447
- Kornberg, A. (1974) DNA Replication (pub. W.H. Freeman)
- Kornberg, A. (1982) DNA Replication Supplement (pub. W.H. Freeman)
- Krigbaum, W.R. & Knutton, S.P. (1973) Proc. Natl. Acad. Sci. U.S.A. 70 2809-2813
- Laemmli, U.K. (1970) Nature 227 680-685
- La Farina, M. & Model, P. (1978) Virology 86 368-375
- Laskey, R.A. & Mills, A.D. (1975) Eur. J. Biochem. 56 335-341
- Lerner, T. & Model, P. (1981) Virology 115 282-294
- Levy, S. & Simpson, R.T. (1973) Nature New Biology 241 139-141
- Lica, L. & Ray, D.S. (1977) J. Molec. Biol. 155 45-59

- Lin, N.S.-C. & Pratt, D. (1972) J. Molec. Biol. 72 37-49
- Lin, T.S., Webster, R.E. & Konigsberg, W. (1980) J. Biol. Chem. 255  
10331-10337
- Loeb, T. (1960) Science 131 932-933
- Lopez, J. & Webster, R.E. (1982) J. Virol. 42 1099-1107
- Lopez, J. & Webster, R.E. (1984) Virology 127 177-193
- Luria S.E. & Burrous J.W. (1957) J. Bacteriol. 74 461-476
- Luria, S.E., Darnell, J.E., Baltimore, D. & Campbell, A. (1978) General  
Virology (pub. J.Wiley & Sons)
- Lyons, L.B. & Zinder, N.D. (1972) Virology 49 45-60
- Maeda, K., Kneale, G.G., Tsugita, A., Short, N.J., Perham, R.N., Hill,  
D. & Petersen, G.B. (1982) E.M.B.O. Jour. 1 255-261
- Makino, S., Woolford, J.L., Tanford, C. & Webster, R.E. (1975) J. Biol.  
Chem. 250 4327-4332
- Makowski, L., Caspar, D.L.D. & Marvin, D.A. (1980) J. Molec. Biol.  
140 149-181
- Mandel, G. & Wickner, W. (1979) Proc. Natl. Acad. Sci. U.S.A. 76  
236-240
- Manning, M., Chrysogelos, S. & Griffith, J. (1981) J. Virol. 40  
912-919
- Marco, R. (1975) Virology 68 280-283
- Marco, R., Jazwinski, S.M. & Kornberg, A. (1974) Virology 62 209-223
- Marvin, D.A. (1966) J. Molec. Biol. 15 8-17
- Marvin, D.A. & Hoffman-Berling, H. (1963) Nature 197 517-518
- Marvin, D.A. & Hohn, B. (1969) Bacteriol. Rev. 33 172-209
- Marvin, D.A., Nave, C., Ladner, J.E., Fowler, A.G., Brown, R.S. &  
Wachtel, E.J. (1981) in Structural Aspects of Recognition and  
Activity in Biological Macromolecules, Ed. Balaban, M., Sussman,  
J.L., Traub, W. & Yonath, A. (pub. Balaban I.S.S.) 819-910

- Marvin, D.A., Pigram, W.J., Wiseman, R.L., Wachtel, E.J. & Marvin, F.J.  
(1974) J. Molec. Biol. 88 581-600
- Marvin, D.A. & Schaller, H. (1966) J. Molec. Biol. 15 1-7
- Marvin, D.A. & Wachtel, E.J. (1975) Nature 253 19-23
- Marvin, D.A. & Wachtel, E.J. (1976) Phil. Trans. Roy. Soc., Lond. B.  
276 81-98
- Marvin, D.A., Wiseman, R.L. & Wachtel, E.J. (1974) J. Molec. Biol. 82  
121-138
- Mazur, B.J. & Zinder, N.D. (1975) Virology 68 490-502
- McHenry, C. & Kornberg, A. (1977) J. Biol. Chem. 252 6478-6484
- McGhee, J.D. & von Hippel, D.H. (1974) J. Molec. Biol. 86 469-489
- McPherson, A., Jurnak, F.A., Wang, A.H.J., Molineux, I. & Rich, A.  
(1979) J. Molec. Biol. 134 378-400
- Merril, C.R., Goldman, D., Sedman, S.A. & Ebert, M.H. (1981) Science  
211 1437-1438
- Meyer, T.F. & Geider, K. (1979a) J. Biol. Chem. 254 12636-12641
- Meyer, T.F. & Geider, K. (1979b) J. Biol. Chem. 254 12642-12646
- Meyer, T.F. & Geider, K. (1981) Proc. Natl. Acad. Sci. U.S.A. 78  
5416-5420
- Meyer, T.F., Geider, K., Kurz, C. & Schaller, H. (1979) Nature 278  
365-367
- Moses, P.B., Boeke, J.D., Horiuchi, K. & Zinder, N.D. (1980) Virology  
104 267-278
- Moses, P.B. & Horiuchi, K. (1982) Virology 119 231-244
- Moses, P.B. & Model, P. (1984) J. Molec. Biol. 172 1-22
- Morrissey, J.H. (1981) Anal. Bioch. 117 307-310
- Nakashima, Y., Dunker, A.K., Marvin, D.A. & Konigsberg, W. (1974a)  
Febs. Letts. 40 290-292
- Nakashima, Y., Dunker, A.K., Marvin, D.A. & Konigsberg, W. (1974b)  
Febs. Letts. 43 125

- Nakashima, Y. & Konigsberg, W. (1974) J. Molec. Biol. 88 598-600
- Nakashima, Y. & Konigsberg, W. (1980) J. Molec. Biol. 138 493-501
- Nakashima, Y., Wiseman, R.L., Konigsberg, W. & Marvin, D.A. (1974c)  
Nature 253 68-71
- Nave, C., Fowler, A.G., Malsey, S., Marvin, D.A. & Seigrist, H. (1979)  
Nature 281 232-234
- Newman, J., Swinney, H.L. & Day, L.A. (1977) J. Molec. Biol. 116  
593-606
- Niyogi, S.K. & Mitra, S. (1977) Biochem. Biophys. Res. Comm. 79  
1037-1044
- Niyogi, S.K. & Mitra, S. (1978) J. Biol. Chem. 253 5563-5567
- Novotny, C., Charnan, J. & Brinton, C.C. (1969a) J. Bacteriol. 98  
1294-1306
- Novotny, C.P. & Fives-Taylor, P. (1974) J. Bacteriol. 117 1306-1311
- Novotny, C., Knight, W.S. & Brinton, C.C. (1968) J. Bacteriol. 95  
314-326
- Novotny, C.P. & Lavin, K. (1971) J. Bacteriol. 107 671-682
- Novotny, C.P., Raizen, E., Knight, W.S., Brinton, C.C. (1969b) J.  
Bacteriol. 98 1307-1319
- Nozaki, Y., Chamberlain, B.K., Webster, R.E. & Tanford, C. (1976)  
Nature 259 335-337
- Nozaki, Y., Reynolds, J.A. & Tanford, C. (1978) Biochemistry 17  
1239-1246
- O'Callaghan, R., Bradley, R. & Paranchych, W. (1973) Virology 54  
220-229
- Ohkawa, I. & Webster, R.E. (1981) J. Biol. Chem. 256 9951-9958
- Ohsumi, M., Vovis, G.F. & Zinder, N.D. (1978) Virology 89 438-449
- Peeters, B.P.H., Konings, R.N.H. & Schoenmakers, J.G.G. (1983) J.  
Molec. Biol. 169 197-215
- Paradiso, P.R. & Konigsberg, W. (1982) J. Biol. Chem. 257 1462-1467

- Paradiso, P.R., Nakashima, Y. & Konigsberg, W. (1979) J. Biol. Chem. 254 4739-4744
- Peeters, B.P.H., Peters, R.M., Schoenmakers, J.G.G. & Konings, R.N.H. (1985) J. Molec. Biol. 181 27-40
- Pratt, D. & Erdahl, W.S. (1968) J. Molec. Biol. 37 181-200
- Pratt, D., Laws, P. & Griffith, J. (1974) J. Molec. Biol. 82 425-439
- Pratt, D., Tzagoloff, H. & Erdahl, H.S. (1966) Virology 30 397-410
- Pretorius, H.T., Klein, M. & Day, L.A. (1975) J. Biol. Chem. 250 262-9269
- Ravetch, J.V., Horiuchi, K. & Zinder, N.D. (1979) J. Molec. Biol. 128 305-318
- Rogers, J., Early, P., Carter, C., Calawe, K., Band, M., Hood, L. & Wall, R. (1980) Cell 20 3030-312
- Rossomando, E.F. & Milstein, J.B. (1971) J. Molec. Biol. 58 187-195
- Rossomando, E.F. & Zinder, N.D. (1968) J. Molec. Biol. 36 387-399
- Rowe, G.W. (1983) J. Theoretical Biol. 101 151-170
- Roy, A. & Mitra, S. (1970a) J. Virol. 6 333-339
- Roy, A. & Mitra, S. (1970b) Nature 228 365-366
- Salivar, W.D., Tzagoloff, H. & Pratt, D. (1964) Virology 24 359-371
- Schaller, H. (1978) Cold Spring Harbor Symp. Quant. Biol. 43 401-408
- Schwartz, F.M. & Zinder, N.D. (1967) Virology 34 352-355
- Shen, C-K.J., Ikoku, A. & Hearst, J.E. (1979) J. Molec. Biol. 127 163-175
- Shimamoto, N. & Utiyama, H. (1983) Biochemistry 22 5869-5878
- Shishido, K. & Ikeda, Y. (1971) J. Molec. Biol. (55) 287-291
- Sigal, N., Delius, H., Kornberg, T., Gefter, M.L. & Alberts, B. (1972) Proc. Natl. Acad. Sci. U.S.A. 69 3537-3541
- Silver, P., Watts, C. & Wickner, W. (1981) Cell 25 341-345
- Simons, G.F.M., Konings, R.N.H. & Schoenmakers, J.G.G. (1979) Febs. Letts. 106 8-12

- Simons, G.F.M., Konings, R.N.H. & Schoenmakers, J.G.G. (1981) Proc. Natl. Acad. Sci. U.S.A. 78 4194-4198
- Skinner, M.K. & Griswold, M.D. (1983) Biochem. J. 209 281-284
- Smilowitz, H. (1974) J. Virol. 13 94
- Smilowitz, H., Carson, J. & Robbins, P.W. (1972) J. Supramolecular Structure 1 8-18
- Smith, H.W. & Huggins, M.B. (1983) J. Gen. Microbiol. 129 2659-2675
- Sugimoto, K., Sugisaki, H., Okamoto, T. & Takanami, M. (1977) J. Molec. Biol. 111 487-507
- Takeya, K. & Amako, K. (1966) Virology 28 163
- Tasheva, B. & Desser, G. (1983) Anal. Biochem. 129 98-102
- Tate, W.P. & Petersen, G.B. (1974a) Virology 57 64-76
- Tate, W.P. & Petersen, G.B. (1974b) Virology 57 77-84
- Tate, W.P. & Petersen, G.B. (1976) Virology 62 17-25
- Thomas, G.J. & Prescott, B. (1983) J. Molec. Biol. 165 321-356
- Torbet, J. (1979) Febs. Letts. 108 61-65
- Torbet, J., Gray, D.M., Gray, G.W., Marvin, D.A. & Siegrist, H. (1981) J. Molec. Biol. 146 305-320
- Tzagoloff, H. & Pratt, D. (1964) Virology 24 372-380
- Van Den Hondel, C.A., Pennings, L. & Schoenmakers, J.G.G. (1976) Eur. J. Biochem. 68 55-70
- Vinuela, E., Algranati, I.D. & Ochoa, S. (1967) Eur. J. Biochem. 1 3-11
- Wachtel, E.J., Marvin, F.J. & Marvin, D.A. (1976) J. Molec. Biol. 107 379-383
- Wachtel, E.J., Wiseman, R.L., Pigram, W.J. & Marvin, D.A. (1974) J. Molec. Biol. 88 601-618
- Watts, C., Silver, P. & Wickner, W. (1981) Cell 25 347-353
- Webster, R.E., Grant, R.A. & Hamilton, L.A.W. (1981) J. Molec. Biol. 152 357-374



- Wickner, W. (1975) Proc. Natl. Acad. Sci. U.S.A. 72 4749-4753
- Wickner, W. (1976) Proc. Natl. Acad. Sci. U.S.A. 73 1159-1163
- Wickner, W. (1977) Biochemistry 16 254-258
- Wickner, W. (1983) T.I.B.S. 8 90-94
- Wickner, W., Brutlag, D., Sheckman, R. & Kornberg, A. (1972) Proc. Natl. Acad. Sci. U.S.A. 69 965-969
- Wickner, W. & Killick, T. (1977) Proc. Natl. Acad. Sci. U.S.A. 74 505-509
- Wickner, W., Mandel, G., Zwizinski, C., Bates, M. & Killick, T. (1978) Proc. Natl. Acad. Sci. U.S.A. 75 1754-1758
- Williams, R.W. & Dunker, A.K. (1977) J. Biol. Chem. 252 6253-6255
- Willams, P.G. & Fenwick, M.L. (1967) Nature 214 712-713
- Wiseman, R.L., Berkowitz, S.A. & Day, L.A. (1976) J. Molec. Biol. 102 549-561
- Wiseman, R.L. & Day, L.A. (1977) J. Molec. Biol. 116 607-611
- Woolford, J.L., Steinman, H.M. & Webster, R.E. (1977) Biochem. J. 16 2694-2700
- Woolford, J.L. & Webster, R.E. (1975) J. Biol. Chem. 250 4333-4339
- Wray, W., Boulikos, T., Wray, V.P. & Hancock, R. (1981) Anal. Biochem. 118 197-203
- Yamamoto, K.R., Alberts, B.M., Benzinger, R., Lawhorne, L. & Treiber, G. (1970) Virology 40 734-744
- Yamamoto, M., Kanegasaki, S. & Yoshikawa, M. (1980) J. Gen. Microbiol. 119 87-93
- Yamamoto, N. & Naito, T. (1965) Science 150 1603-1604
- Zentgraf, H., Berthold, V. & Geider, K. (1977) Biochim. Biophys. Acta 474 629-638
- Zinder, N.D., Valentine, R.C., Roger, M. & Stockenius, W. (1963) Virology 20 638-640

Zlochevski, M.L., Kozlov, Y.I., Tal, E.K. & Debabov, V.G. (1977)

Molekularnya Biologiya 11 217-227

Zwizinski, C. & Wickner, W. (1977) Biochim. Biophys. Acta 471 169-176

## Appendix I - Useful Figures

For phage fd:

Length: 895 nm	Diameter: 5.5 nm	Newman <u>et al.</u> , 1977
$\frac{M}{r}$ 16.4 x 10 <sup>6</sup>		Newman <u>et al.</u> , 1977
A <sub>269</sub> (1 mg/ml, 1 cm) = 3.84		Nozaki <u>et al.</u> , 1976
DNA length: 6408 nucleotides		Beck <u>et al.</u> , 1978
No. of g8p subunits: 2700		Newman <u>et al.</u> , 1977
Ratio subunits/nucleotides: 2.3:1		Day <u>et al.</u> , 1979

For phage Pf1:

Length: 1960 nm	Diameter: 4.5 nm	Torbet, 1979
$\frac{M}{r}$ 38.4 x 10 <sup>6</sup>		Wiseman & Day, 1976
A <sub>270</sub> (1 mg/ml, 1 cm) = 2.07		Nozaki <u>et al.</u> , 1976
DNA length: 7690 nucleotides		N.J. Short, pers. comm.
No. of g8p subunits: 7620		Wiseman & Day, 1977
Ratio nucleotides/subunits: 1:1		Day <u>et al.</u> , 1979

Appendix II - Fd and Pf1 Open Reading Frames

Gene	fd: $\frac{M}{-r}$	Polarity Index	Pf1: $\frac{M}{-r}$	Polarity Index
I	39,536	47	38,706	39
II	46,235	48	11,168	36
III	44,605	43	44,959	43
IV	45,821	49	49,103	45
V	9,688	43	15,428	44
VI	12,349	26	15,773	20
VII	3,602	30	3,217	35
VIII	7,626	34	8,377	32
IX	3,653	38	8,644	27
X	12,680	45	12,141	50

Figures obtained by application of program "MW-calc"  
(Dr. P.R. Alefounder) to the sequence data for fd (Beck  
et al., 1978), and Pf1 (N.J. Short, pers. comm.).

```

>LIST
10MODE 3
20PRINT '***'.
*** PROTEIN FOLD ***
... by FRANK CSELIK .

30PRINT "THIS PROGRAM PREDICTS PROTEIN SECONDARY STRUCTURE FROM AMINO ACID SE-
QUENCES." "THE SEQUENCES MUST BE FILED ON DISC USING THIS PROGRAM." "A DERIVAT-
IVE OF CHOU-FASMAN ANALYSIS IS USED GIVING OUTPUT IN SEVEN LINES : "
40PRINT TAB(9);": THE FRAME POSITION MARKER, EVERY 10 RESIDUES"
50PRINT "DLLAVGGKPM THE AMINO ACID SEQUENCE IN SINGLE LETTER CODE"
60PRINT " !! ! ALPHA HELIX INHIBITION ! > * "
70PRINT "oooo- ALPHA HELIX POTENTIAL 0 > o > - ) ASSIGNED AT THE " ' T-
AB(45);"> CENTRE OF EACH"
80PRINT "vvvv-- - BETA SHEET POTENTIAL V > v > - ) PENTA-RESIDUE"
90PRINT " ***! BETA SHEET INHIBITION ! > * "
110PRINT "-.... +.. POLARITY : HYDROPHOBIC (.), HYDROPHILIC ( ), CHARGED (+
) OR (-)"
120INPUT "DO YOU WISH TO FILE A SEQUENCE ? "R$
130PRINT
140IF R$<>"Y" THEN 200
150INPUT "PLEASE GIVE THE SEQUENCE A NAME : "A$
160IF LEFT$(A$,1)="X" THEN 180 ELSE PRINT "SUITABLE FILE NAMES BEGIN WITH THE
LETTER X" : INPUT "WOULD YOU PLEASE GIVE THE SEQUENCE A SUITABLE NAME "A$
170IF LEFT$(A$,1)="X" THEN 180 ELSE PRINT "OK ! BYE-BYE !" : GOTO 1820
180Y=OPENOUT A$ : PRINT "PLEASE ENTER UP TO 70 RESIDUES OF SEQUENCE IN SINGLE
LETTER CODE." TAB(0);"[";TAB(71);"]" : INPUT P$ : IF LEN(P$)>70 THEN P$=LEFT$(P
$,70) : PRINT "SEQUENCE TRUNCATED TO 70 RESIDUES."
190PRINT#Y,P$ : CLOSE #Y : INPUT "DO YOU WISH TO FILE ANOTHER SEQUENCE ? "R$
: IF R$="Y" THEN PRINT : GOTO 140
200 DIM H(70), B(70)
210FOR I=1 TO 70 : H(I)=0 : B(I)=0 : NEXT
220A$="O"
230I=1 : C=0 : D=0
240INPUT "PLEASE GIVE THE NAME OF THE SEQUENCE TO BE ANALYSED " K$
250IF LEFT$(K$,1)="X" THEN 320
260PRINT "SUITABLE FILE-NAMES BEGIN WITH THE LETTER X"
270INPUT "PLEASE GIVE THE NAME OF A FILE SUITABLE FOR THIS ANALYSIS "K$
280IF LEFT$(K$,1)="X" THEN 320
290PRINT "BYEEEEEE...."
300STOP
310REM ** ASSIGN PARAMETERS **
320X=OPENIN K$
330INPUT #X, P$
340G$=MID$(P$,1,1)
350IF I>70 THEN 630
360IF G$="E" THEN C=1.53 : D=0.26
370IF G$="A" THEN C=1.45 : D=0.97
380IF G$="L" THEN C=1.34 : D=1.22
390IF G$="H" THEN C=1.24 : D=0.71
400IF G$="M" THEN C=1.20 : D=1.67
410IF G$="Q" THEN C=1.17 : D=1.23
420IF G$="W" THEN C=1.14 : D=1.19
430IF G$="V" THEN C=1.14 : D=1.65
440IF G$="F" THEN C=1.12 : D=1.28
450IF G$="K" THEN C=1.07 : D=0.74
460IF G$="I" THEN C=1.00 : D=1.60
470IF G$="D" THEN C=0.98 : D=0.80
480IF G$="T" THEN C=0.82 : D=1.20
490IF G$="S" THEN C=0.79 : D=0.72
500IF G$="R" THEN C=0.79 : D=0.90
510IF G$="C" THEN C=0.77 : D=1.30
520IF G$="N" THEN C=0.73 : D=0.65
530IF G$="Y" THEN C=0.61 : D=1.29
540IF G$="P" THEN C=0.59 : D=0.62
550IF G$="G" THEN C=0.53 : D=0.81
560 IF C=0 OR D=0 THEN 580
570GOTO 600
580PRINT "TRY AGAIN, PLEASE."
590 GOTO 230
600H(I)=C : B(I)=D : A$=A$+G$ : I=I+1
610IF I=LEN(P$)+1 THEN 630
620 GOTO 340
630LET N=LEN(A$)-1
640REM ** PRINT MARKERS **
650PRINT'
660FOR I=1 TO 7 : PRINT TAB(I*10-1);":": : NEXT
670PRINT
680REM ** PRINT AMINO ACID SEQUENCE **
690FOR J=1 TO 70
700IF J>N THEN 730
710 PRINT MID$(A$,(J+1),1);
720GOTO 750
730REM
740PRINT " ";
750NEXT J
760PRINT
770GOTO 1220
780REM ** ASSIGN HELIX POTENTIAL **
790PRINT "..."
800 FOR I=3 TO (N-2)
820LET C=0
830FOR J=0 TO 4
840 LET Q=I+J-2
850LET C=C+H(Q)
860NEXT J
870LET F$=" "
880IF C>=4.95 THEN F$="-"
890IF C>=5.475 THEN F$="o"
900IF C>=6.50 THEN F$="O"

```

```

800FOR I=3 TO (N-2)
820LET C=0
830FOR J=0 TO 4
840LET Q=I+J-2
850LET C=C+H(Q)
860NEXT J
870LET F$=" "
880IF C>=4.95 THEN F$="-"
890IF C>=5.475 THEN F$="o"
900IF C>=6.50 THEN F$="Q"
910GOTO 930
920LET F$="I"
930PRINT F$;
940NEXT I
950PRINT "..."
960REM ** ASSIGN SHEET POTENTIAL **
970PRINT "...";
980FOR I=3 TO (N-2)
990LET C=0
1000FOR J=0 TO 4
1010LET Q=I+J-2
1020LET C=C+B(Q)
1030NEXT J
1040LET F$=" "
1050IF C>=4.675 THEN F$="-"
1060IF C>=5.55 THEN F$="v"
1070IF C>=7.25 THEN F$="V"
1080PRINT F$;
1090NEXT I
1100PRINT "..."
1110GOTO 1430
1220REM ** ASSIGN HELIX STOPS **
1230FOR I=1 TO N
1240F$=" "
1250IF I>2 THEN 1260 ELSE 1280
1260IF H(I)<0.85 AND H(I-1)<0.85 AND H(I-2)<0.85 THEN F$="*"
1270IF H(I)<0.6 AND F$="*" THEN F$="!"
1280IF I<(N-1) THEN 1290 ELSE 1310
1290IF H(I)<0.85 AND H(I+1)<0.85 AND H(I+2)<0.85 THEN F$="*"
1300IF H(I)<0.6 AND F$="*" THEN F$="!"
1310IF I>1 THEN 1320 ELSE 1330
1320IF H(I)<0.6 AND H(I-1)<0.6 THEN F$="!"
1330IF I<N THEN 1340 ELSE 1350
1340IF H(I)<0.6 AND H(I+1)<0.6 THEN F$="!"
1350IF I<N AND I>1 THEN 1360 ELSE 1380
1360IF H(I)<0.85 AND H(I+1)<0.85 AND H(I-1)<0.85 THEN F$="*"
1370IF H(I)<0.6 AND F$="*" THEN F$="!"
1380IF H(I)=0.59 THEN F$="!"
1385PRINT F$;
1390NEXT I
1400PRINT
1410GOTO 780
1420REM ** ASSIGN SHEET STOPS **
1430FOR I=1 TO N
1440F$=" "
1450IF I>2 THEN 1460 ELSE 1470
1460IF B(I)<0.94 AND B(I-1)<0.94 AND B(I-2)<0.94 THEN F$="*"
1465IF B(I)<0.73 AND F$="*" THEN F$="!"
1470IF I<(N-1) THEN 1480 ELSE 1490
1480IF B(I)<0.94 AND B(I+1)<0.94 AND B(I+2)<0.94 THEN F$="*"
1485IF B(I)<0.73 AND F$="*" THEN F$="!"
1490IF I>1 THEN 1500 ELSE 1510
1500IF B(I)<0.73 AND B(I-1)<0.73 THEN F$="!"
1510IF I<N THEN 1520 ELSE 1530
1520IF B(I)<0.73 AND B(I+1)<0.73 THEN F$="!"
1530IF I<N AND I>1 THEN 1540 ELSE 1550
1540IF B(I)<0.94 AND B(I+1)<0.94 AND B(I-1)<0.94 THEN F$="*"
1545IF B(I)<0.73 AND F$="*" THEN F$="!"
1550PRINT F$;
1560NEXT I
1570PRINT
1590REM ** ASSIGN POLARITIES **
1600FOR I=2 TO (N+1)
1610LET E$=MID$(A$,I,1)
1620IF E$="K" OR E$="R" OR E$="H" THEN PRINT TAB(I-2); "+";
1630IF E$="D" OR E$="E" THEN PRINT TAB(I-2); "-";
1640IF E$="A" OR E$="L" OR E$="I" OR E$="M" OR E$="P" OR E$="F" OR E$="V" THEN PRINT TAB(I-2); ".";
1650NEXT I
1660PRINT ""
1670CLOSE EX
1680INPUT "DO YOU REQUIRE ANOTHER ANALYSIS ? "R$
1690PRINT
1700IF R$="Y" THEN 210
1710INPUT "DO YOU WISH TO FILE A SEQUENCE ? "R$
1720IF R$="Y" THEN 150
1730INPUT "DO YOU REQUIRE ADVICE ON THE INTERPRETATION OF OUTPUT ? "R$
1740IF R$<>"Y" THEN 1820
1750PRINT ""ALPHA HELICES AND BETA PLEATS PROPAGATE BI-DIRECTIONALLY FROM INITIATION ZONES""(oooo) OR (vvvv) ALONG THE POLYPEPTIDE. THEY STOP AT INHIBITION SEQUENCES (|)|""OR (|*|) OR (|*|) OR (**|) OR (***).""
1760PRINT "P D E ARE STABILIZING AT THE N-TERMINAL END OF AN ALPHA HELIX REGION."
1770PRINT "H K R ARE STABILIZING AT THE C-TERMINAL END OF AN ALPHA HELIX REGION."
1770PRINT "THE OUTPUT SHOWS RELATIVE STABILIZATIONS OF THE COMPETING STRUCTURES."
1780PRINT "THE STRUCTURE OF CONSISTENTLY HIGHER PROBABILITY IN A REGION SHOULD BE ASSIGNED."
1780PRINT "SOME SEQUENCES - ESPECIALLY OF PHAGE PROTEINS - CAN ADOPT EITHER CONFORMATION."
1790PRINT "THIS IS SHOWN QUITE CLEARLY IN THE OUTPUT AND IS NOT USUALLY THE RESULT OF "
1790PRINT "AMBIGUITY. RUNS OF (oooo) AND (vvvv) SUGGEST

```



```

1650 IF E$="A" OR E$="L" OR E$="I" OR E$="M" OR E$="P" OR E$="F" OR E$="V" TH
EN PRINT TAB(I-2); ". ";
1650 NEXT I
1660 PRINT ""
1670 CLOSE EX
1680 INPUT "DO YOU REQUIRE ANOTHER ANALYSIS ? "R$
1690 PRINT
1700 IF R$="Y" THEN 210
1710 INPUT "DO YOU WISH TO FILE A SEQUENCE ? "R$
1720 IF R$="Y" THEN 150
1730 INPUT "DO YOU REQUIRE ADVICE ON THE INTERPRETATION OF OUTPUT ? "R$
1740 IF R$<>"Y" THEN 1820
1750 PRINT ""ALPHA HELICES AND BETA PLEATS PROPAGATE BI-DIRECTIONALLY FROM INIT
IATION ZONES""(oooo) OR (vvvv) ALONG THE POLYPEPTIDE. THEY STOP AT INHIBITION S
EQUENCES (|)|""OR (*|*) OR (|*|) OR (**|) OR (***).""
1760 PRINT "P D E ARE STABILIZING AT THE N-TERMINAL END OF AN ALPHA HELIX REGION
.""H K R ARE STABILIZING AT THE C-TERMINAL END OF AN ALPHA HELIX REGION.""
1770 PRINT "THE OUTPUT SHOWS RELATIVE STABILIZATIONS OF THE COMPETING STRUCTURES
.""THE STRUCTURE OF CONSISTENTLY HIGHER PROBABILITY IN A REGION SHOULD BE ASSIG
NED.""
1780 PRINT "SOME SEQUENCES - ESPECIALLY OF PHAGE PROTEINS - CAN ADOPT EITHER CON
FORMATION.""THIS IS SHOWN QUITE CLEARLY IN THE OUTPUT AND IS NOT USUALLY THE RE
SULT OF "
1790 PRINT "AMBIGUITY. RUNS OF (oooo) AND (vvvv) SUGGEST ALTERNATIVE STRUCTURES
THAT MAY ""BE STABILIZED ACCORDING TO ENVIRONMENT.""
1800 PRINT "POSSIBLE MEMBRANE INSERTION SEQUENCES ARE ALSO DISPLAYED AS RUNS OF
(.....) ""AND ( ) IN THE POLARITY PROFILE WITH NO CHARGES BUT FLANKED BY CHAR
GED""RESIDUES ACTING AS ANCHORS.""
1810 GOTO 1680
1820 PRINT "BYE-BYE"
1830 END

```